



**HAL**  
open science

# Outils et méthodes pour le codage vidéo perceptuel

Vincent Ricordel

► **To cite this version:**

Vincent Ricordel. Outils et méthodes pour le codage vidéo perceptuel. Traitement du signal et de l'image [eess.SP]. Université de Nantes, 2020. tel-02461613

**HAL Id: tel-02461613**

**<https://theses.hal.science/tel-02461613v1>**

Submitted on 30 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Nantes  
Laboratoire des Sciences du Numérique de Nantes (LS2N, UMR 6004 CNRS)  
Ecole Polytechnique de l'Université de Nantes

# Outils et méthodes pour le codage vidéo perceptuel

Mémoire présenté par  
**Vincent Ricordel**  
en vue d'obtenir l'**Habilitation à Diriger des Recherches**

Ecole doctorale Math-STIC  
(Mathématiques et Sciences et Technologies de l'Information et de la Communication)  
Spécialité : Signal, Image, Vision

Date de soutenance : 22 janvier 2020

---

Rapporteurs et membres du jury :

- Marc Antonini, DR. à l'Université de Nice
- William Puech, Pr. à l'Université de Montpellier
- Olivier Déforges, Pr. à l'INSA de Rennes

Membres du jury :

- Marco Cagnazzo, Pr. à l'Institut Télécom-Paris
  - Luce Morin, Pr. à l'INSA de Rennes
  - Marc Gelgon, Pr. à l'Université de Nantes
  - Bogdan Cramariuc, Directeur de l'ITCST à Bucarest
-



# Remerciements

Je tiens à exprimer tous mes remerciements.

A Marc Antonini, William Puech et Olivier Déforges qui ont accepté d'être les rapporteurs de ce manuscrit et de participer à mon jury. Ainsi qu'à Marco Cagnazzo, Luce Morin, Marc Gelgon et Bogdan Cramariuc, qui ont bien voulu s'associer à mon jury. C'est toujours un plaisir de retrouver ces collègues que je connais depuis longtemps (certains depuis plus de 20 ans... ). Leurs qualités scientifiques sont indéniables, et ils sont devenus de véritables compagnons de route m'accompagnant tout au long de ma carrière.

A tous ceux qui ont contribué à ces travaux. Je pense en particulier aux nombreux jeunes que j'ai eu la chance d'encadrer durant leur stage de master, ou leur thèse de doctorat. Comme le détaille ce manuscrit, il y a eu successivement les thèses de Guillaume Jeannic, Olivier Brouard, Junle Wang, et Karam Naser (sachant que celle d'Amira Filali est en cours). Chacun me laisse un souvenir fort, et comme j'ai régulièrement de leurs nouvelles, je sais que chacun a trouvé sa voie et la trace solidement. J'en profite pour remercier Dominique Barba et Patrick Le Callet avec qui j'ai co-encadré ces thèses. La passion et tous les savoirs qu'ils m'ont montré dans leurs activités de recherche étaient remarquables, et j'aimerais être capable de suivre leur exemple.

Ces travaux ont été effectués au sein des équipes IVC de l'IRCCyN puis IPI du LS2N, équipes elles-mêmes hébergées physiquement par l'Ecole Polytechnique de l'Université de Nantes. Je souhaite également remercier profondément l'ensemble des collègues (et je ne peux pas tous les citer), qui ont fait de ce lieu un endroit unique où se développent conjointement une ambiance chaleureuse, ainsi qu'un enseignement et une recherche de qualité.

Enfin, mes pensées vont à mes proches, qui toujours me soutiennent, et grâce auxquels je continue de grandir.



# Résumé

On ne peut que constater combien la technologie vidéo est devenue omniprésente en un peu plus d'une décennie. Si les besoins des nouveaux services vidéo ont été en partie satisfaits par l'efficacité des réseaux et de leurs technologies de couche physique, en parallèle les codecs vidéo ont été primordiaux en offrant compression et nouvelles fonctionnalités. La recherche en codage vidéo se caractérise alors par un fort contexte normatif où le point de compatibilité se fait au niveau du décodage. Les performances de chaque nouveau standard sont remarquables, car typiquement chacun par rapport au précédent, code avec la même qualité objective un format vidéo donné mais en divisant par deux le débit nécessaire. Il est cependant frappant de constater que le paradigme de la couche de codage n'a pas changé depuis l'origine, avec un codec dissymétrique où les décisions sont prises à l'encodage via un schéma hybride intégrant les étapes de prédiction, transformation, quantification et codage entropique. Il est aussi surprenant de constater la persistance de l'implantation dans les codeurs pour la phase d'optimisation débit-distorsion, d'un calcul de distorsion de type erreur quadratique moyenne, alors que l'on sait la corrélation pauvre entre les résultats de cette métrique et le jugement humain. C'est à ce niveau que se définit l'objectif du codage vidéo perceptuel qui est d'aller vers une représentation de l'information basée sur le contenu perçu et qui puisse servir de paradigme pour un codage basé qualité perceptuelle des contenus des vidéos 2D et 3D. Pour être en phase avec les approches standards, l'objectif devient de guider les choix de codage de façon à aller vers la fidélité perceptuelle sans perte des vidéos décodées. Mes travaux se positionnent donc dans cette recherche des outils et des méthodes où les choix pour améliorer la qualité perceptuelle guident le codage vidéo.

Mon mémoire retrace mon itinéraire scientifique au sein des équipes IVC IRCCyN puis IPI LS2N qui, du fait des recherches qui y sont conduites sur la modélisation du système visuel humain, ont offert et offrent encore un cadre collaboratif particulièrement propice aux travaux liés au codage vidéo perceptuel. Mes contributions au codage vidéo perceptuel se sont principalement organisées au sein de projets de recherche visant à contribuer aux efforts faits en normalisation, avec le projet ANR ArchiPEG qui avait pour objectif l'optimisation du codage vidéo HD en flux H.264/AVC, le projet ANR PERSEE qui s'intéressait au codage perceptuel des formats vidéo 2D et 3D via le codec H.264 et son extension MVC, et le projet européen ITN PROVISION qui visait le codage perceptuel via HEVC. Mes contributions ont été produites au travers l'encadrement de masters et de thèses (4 soutenues, 1 en cours), et nous verrons que nous avons agi à 2 niveaux de

la boucle de codage vidéo hybride.

Une première possibilité est d'améliorer directement des unités de la boucle de codage. C'est ainsi que nous avons mis en œuvre une transformation directionnelle tenant compte des orientations géométriques au sein de l'image. Nous avons exploité cette représentation de l'information afin de construire des classes de structures, puis adapté la quantification de leur contenu en se basant sur des critères perceptuels. Nous avons aussi, après avoir défini la notion de textures spatio-temporelles visuelles, proposé un modèle mesurant la similarité entre ces textures visuelles. Ce modèle computationnel basé sur le traitement cortical du SVH, a ensuite été utilisé afin d'améliorer la boucle de codage de HEVC : en identifiant des classes de textures visuelles et en faisant correspondre à la distorsion objective standard calculée par le codeur sa valeur dans le domaine perceptuel.

Une autre approche consiste à mettre en place un schéma de pré-analyse avant encodage afin de piloter ensuite efficacement le codage des régions d'intérêt. La modélisation de l'attention visuelle est donc essentielle au niveau de cette analyse. Nous avons conçu un schéma de pré-analyse basé sur des tubes spatio-temporels. Cet outil réalise la segmentation et le suivi des objets constitués de tubes connexes, et il leur associe une saillance visuelle. L'ensemble de ces informations est exploitée afin d'assurer un codage cohérent des objets en flux H.264/AVC. Exactement nous avons testé l'adaptation de la structure du GOP, et le guidage de la qualité visuelle. Nous avons aussi développé le premier modèle computationnel de l'attention visuelle 3D s'appliquant à la visualisation d'images stéréoscopiques. Ce modèle combine deux cartes de saillance visuelle : celle 2D et celle de profondeur. Nous avons utilisé ce modèle afin d'améliorer la qualité d'expérience de visualisation de la TV3D, par ajout d'un flou de défocalisation sur la région ayant moins d'intérêt.

Dans la dernière partie de mon mémoire, un bilan et des perspectives sont donnés. Nous nous intéressons en particulier au codage des formats vidéo immersifs représentés par des nuages de points 3D. Nos derniers travaux montrent l'intérêt d'une approche basée quantification vectorielle algébrique et arborescente, permettant de structurer graduellement la géométrie de la scène, et cela en fonction d'un critère débit-distorsion.

**Mots-clés :**

Codage vidéo, Codage vidéo perceptuel, H.264/AVC, H.265/HEVC, HVS, Système visuel humain, Projet ANR ArchiPEG, Projet ANR PERSEE, Projet ITN PROVISION, Transformation en ondelettes orientées, Textures visuelles, Similarité entre textures, Tube spatio-temporel, Attention visuelle, Attention visuelle 3D, Nuages de points 3D, Quantification vectorielle algébrique et arborescente.

# Table des matières

<b>1</b>	<b>Curriculum Vitæ</b>	<b>1</b>
1.1	Identification . . . . .	1
1.2	Diplômes . . . . .	1
1.3	Déroulement de carrière . . . . .	2
1.4	Activités pédagogiques . . . . .	2
1.4.1	Responsabilités collectives liées à la pédagogie . . . . .	5
1.5	Activités en Recherche . . . . .	6
1.5.1	Encadrements de thèses et de masters . . . . .	7
	Encadrement doctoral . . . . .	7
	Encadrement de masters de recherche . . . . .	8
	Encadrement de mémoires CNAM . . . . .	9
1.5.2	Autres activités scientifiques . . . . .	9
	Responsabilités scientifiques . . . . .	10
	Participations à des projets . . . . .	10
	Collaborations avec des chercheurs invités . . . . .	11
	Expertises et collaborations avec l'environnement socio-économique . . . . .	11
	Membre de Conseils et de Comités . . . . .	12
	Animations de la recherche . . . . .	12
1.5.3	Publications . . . . .	14
<b>2</b>	<b>Introduction générale</b>	<b>20</b>
2.1	Positionnement du travail . . . . .	20
2.2	Contributions et annonce du plan . . . . .	24



<b>I</b>	<b>Contributions aux projets de recherche</b>	<b>27</b>
	Introduction . . . . .	28
<b>3</b>	<b>Le projet ANR ArchiPEG</b>	<b>30</b>
	3.1 Contexte et objectifs . . . . .	30
	3.2 Organisation . . . . .	31
	3.3 Travaux menés et résultats . . . . .	32
	3.4 Contributions et bilan personnels . . . . .	35
<b>4</b>	<b>Le projet ANR PERSEE</b>	<b>37</b>
	4.1 Contexte et objectifs . . . . .	37
	4.2 Organisation . . . . .	38
	4.3 Travaux menés et résultats . . . . .	40
	4.4 Contributions et bilan personnels . . . . .	44
<b>5</b>	<b>Le projet ITN PROVISION</b>	<b>46</b>
	5.1 Contexte et objectifs . . . . .	46
	5.2 Organisation . . . . .	47
	5.3 Contributions et bilan personnel . . . . .	50
<b>II</b>	<b>Contributions aux travaux de thèses</b>	<b>52</b>
	Introduction . . . . .	53
<b>6</b>	<b>Représentation d’images par transformées locales orientées et classes structurelles</b>	<b>54</b>
	6.1 Problématique . . . . .	54
	6.2 Travaux menés et résultats . . . . .	56
	6.2.1 Etude de techniques de lifting orienté . . . . .	56
	6.2.2 Estimation de la géométrie de l’image . . . . .	59
	6.2.3 Codage de la géométrie de l’image . . . . .	61
	6.2.4 Quantification perceptuelle des coefficients d’ondelettes orientées . . . . .	63
	6.3 Discussion . . . . .	64

<b>7</b>	<b>Outil de pré-analyse avant encodage vidéo perceptuel</b>	<b>69</b>
7.1	Problématique	69
7.2	Travaux menés et résultats	73
7.2.1	Segmentation basée mouvement	73
7.2.2	Segmentation spatio-temporelle et suivi des objets	76
7.2.3	Carte de saillance visuelle	81
7.2.4	Applications au codage vidéo perceptuel	84
7.3	Discussion	88
<b>8</b>	<b>Modélisation computationnelle de l'attention visuelle 2D &amp; 3D</b>	<b>90</b>
8.1	Problématique	90
8.2	Travaux menés et résultats	91
8.2.1	Comparaisons de vérités de terrain de l'attention visuelle 2D	91
8.2.2	Modélisation computationnelle de l'attention visuelle 3D	93
8.2.3	Amélioration de la qualité d'expérience de visualisation de la TV3D	96
8.3	Discussion	97
<b>9</b>	<b>Modélisation de la similarité des textures visuelles, et compression</b>	<b>106</b>
9.1	Problématique	106
9.2	Travaux menés et résultats	106
9.2.1	Définition et modélisation des textures visuelles	106
9.2.2	Similarité visuelle entre textures, et compression	107
9.2.3	Métrique de similarité V1-E entre textures visuelles	109
9.2.4	Application à la compression vidéo perceptuelle	112
9.3	Discussion	117
<b>10</b>	<b>Représentation efficace des contenus immersifs 3D</b>	<b>121</b>
10.1	Contexte et problématique	121
10.2	Travaux menés et premiers résultats	123
10.2.1	Approches de découpage	123
10.2.2	Métriques et optimisation débit-distorsion	125
10.2.3	Premiers résultats	126
10.3	Futurs travaux	128

<b>11 Conclusion générale et perspectives</b>	<b>131</b>
<b>Table des figures</b>	<b>135</b>
<b>Index</b>	<b>140</b>
<b>Bibliographie</b>	<b>194</b>

# Chapitre 1

## Curriculum Vitæ

### 1.1 Identification

**Nom** Ricordel

**Prénom** Vincent

**Date de naissance** 9 août 1967

**Situation** marié, 3 enfants

**Grade** MCF, hors classe, 5<sup>e</sup> échelon

**Etablissement** Université de Nantes (Polytech Nantes, département informatique)

**Section CNU** 27 (rejointe en 2005)

**Unité de recherche d'appartenance** Laboratoire LS2N (Laboratoire des Sciences du Numérique de Nantes, UMR 6004) dans l'équipe IPI (Image Perception Interaction)

**Site web** [www.pagesperso.ls2n.fr/~ricordel](http://www.pagesperso.ls2n.fr/~ricordel)

### 1.2 Diplômes

J'ai obtenu tous mes diplômes à l'Université de Rennes I.

**1996, Thèse de Doctorat** mention Traitement du Signal et Télécommunication

**Titre** « *Étude de schémas de quantification vectorielle algébrique et arborescente. Application à la compression de séquences d'images numériques* »

**Laboratoire** IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires), projet TEMIS (Traitement, Exploitation, Modélisation d'Images Séquentielles)

**Jury** René Collorec (président), Jean-Pierre Asselin de Beauville (rapporteur), Michel Barlaud (rapporteur), Odile Macchi (examinatrice), Jean-Paul Guillois (examinateur), Claude Labit (directeur), Christine Guillemot (invité)

**1992, DEA STIR** (Signal, Télécommunications, Image, Radar), filière Image

**1991, Maîtrise EEA** (Electronique, Electrotechnique, Automatique)

**1989, DUT GEII** (Génie Electrique et Informatique Industrielle)

### 1.3 Déroulement de carrière

**Depuis 2002** Maître de conférences – après mutation - à l’Université de Nantes : Enseignement à Polytech Nantes et recherche dans le laboratoire IRCCyN, puis celui LS2N.

**1997-2002** Maître de conférences - stage puis titularisation - à l’Université de Toulon et du Var : Enseignement à l’IUT et recherche dans le laboratoire SIS (Systèmes, Information et Signal)

**1996-1997** Séjour postdoctoral au Laboratoire SPL (Signal Processing Laboratory) de l’Université de Technologie de Tampere (TUT, Finlande)

### 1.4 Activités pédagogiques

De 1997 à 2002 à Toulon, j’ai effectué l’essentiel de ma charge à l’IUT pour le département SRC (Services et Réseaux de Communication), cette spécialité est devenue à présent celle MMI (Métiers du Multimédia et de l’Internet). Ce département préparait en 2 ans à un diplôme de technicien supérieur spécialiste des réseaux et des nouvelles techniques de communication. Mes interventions s’inscrivaient dans le pôle scientifique de la formation, et visaient à donner aux étudiants le bagage nécessaire pour appréhender les aspects technologiques abordés au cours de leur cursus à l’IUT. J’étais alors responsable de 4 modules d’enseignement (maths appliquées, électronique, réseaux informatiques, normes de compression) dont j’ai conçu et organisé les cours.

Depuis 2002 à Nantes j’enseigne essentiellement à Polytech, pour les spécialités Informatique et ETN (Electronique et Technologies Numériques). L’enseignement y est semestrialisé sur les deux années du cycle préparatoire (PEIP semestres 1 à 4) et les trois années du cycle ingénieur (semestres 5 à 10). Ma charge d’enseignement oscille autour de 300h, son volume est ventilé en CM (cours magistraux), TD (travaux dirigés), TP (travaux pratiques) et Pr (projets). Les matières que j’enseigne sont majoritairement dans le domaine informatique (réseaux informatiques, développement de services, maths appliquées aux réseaux ou au multimédia), et aussi liées à mes activités de recherche (traitement et compression des images).

J’effectue aussi des interventions courtes en Master Informatique, et uniquement pour des cours liés à ma spécialité de recherche. Je suis responsable de six modules d’enseignement à Polytech. J’ai particulièrement contribué à la conception de cours relatifs au multimédia, aux services et réseaux informatiques, pour les départements Informatique et ETN à Polytech, et pour des Masters.

Mes activités pédagogiques sont présentées synthétiquement ci-après, et le tableau 1.1 résume mon enseignement actuel à Polytech Nantes :

**Depuis 2002** à l'Université de Nantes

- Formations :
  - École Polytechnique de l'université de Nantes, départements Informatique et ETN (Electronique et Technologies Numériques), et cycle préparatoire PEIP
  - Masters :
    - Depuis 2017 : master en informatique parcours international VICO (VIsual COmputing)
    - 2013-2017 : master international MDM (Multimedia and Data Management)
    - 2006-2010 : master M2R SEGE (Systèmes Electronique et Génie Electrique)
    - 2004-2008 : master international hTech (Sciences et Technologies en Informatique et Communication)
    - 2002-2008 : master M2R ALD (Architectures Logicielles Distribuées)
  - Formation continue :
    - 2015-2018 : DUTIL (certification professionnelle de niveau II, Chef de projet Logiciel et Réseaux)
    - 2006-2010 : CNAM (Conservatoire National des Arts et Métiers) cycle ingénieur
- Matières enseignées : Réseaux informatiques, maths appliquées, multimédia (traitement des images et de la vidéo, compression)
- Responsables de 6 modules d'enseignement à Polytech Nantes et co-responsable d'un pour le master en informatique parcours VICO
- Encadrement de stages et de projets (ingénieurs et Master recherche), suivi d'alternants

**1997 à 2002** à l'Université de Toulon et du Var

- Formations :
  - IUT département SRC (Services et Réseaux de Communication)
  - Maîtrise EEA (Électronique Électrotechnique et Automatique)
  - Diplôme d'études Supérieures Universitaire (DESU SML - formation continue)
- Matières enseignées : Réseaux, mathématiques appliquées, compression des images, électronique
- Encadrement de stages et de projets (techniciens, ingénieurs, recherche), et suivi d'apprentis
- Responsable de 4 modules d'enseignement

Semestre	Désignation	CM	TD	TP	Projet
2	Projet Peip				20
5	Info - Réseaux 1 (Principes des Réseaux Informatiques)	14,5	6	4	
6	Info - Analyse de Fourier Info - Réseaux 2 (Conception de protocoles de Communication)		12 13,5	3	
7	Info - Traitement d'image Info - Projet transversal Info - Traitement du Signal ETN - Signaux Multimédia (traitement des images)	7,5 5	7,5 9	12 9 10,5	4,5
8	Info - Projet transversal Info - Données et systèmes Multimédias ETN - Réseaux informatiques ETN - Projet transversal	6,25	9	10,5 12	12 15
9	Info - Projet recherche et développement ETN - Technologies et réseaux multimédia ETN - Technologies et normes multimédia ETN - Communications et services multimédias ETN - Projet recherche et développement	7,5 3		15 12	12 10
10	Info - Projet de fin d'études ETN - Projet de fin d'études				15 15

TABLE 1.1 – Mon enseignement actuel à Polytech Nantes (avec les volumes horaires).

### 1.4.1 Responsabilités collectives liées à la pédagogie

De 1997 à 2002, j'ai assuré de nombreuses responsabilités collectives à l'université de Toulon et du Var, et plus particulièrement en devenant en septembre 2000 le Directeur du département SRC de l'IUT. Ce département était physiquement délocalisé à St Raphaël. Il était constitué d'une formation initiale (2 promotions de 85 étudiants chacune), et d'une formation par alternance créée en 2000 (2 promotions de 24 étudiants chacune). L'équipe pédagogique était constituée d'une cinquantaine d'enseignants (10 titulaires, 5 ATER, 40 vacataires dont de nombreux professionnels) et de personnel IATOS (3 techniciens, 2 secrétaires, 2 contrats aidés).

Depuis 2002 à l'université de Nantes, j'ai continué à prendre des responsabilités collectives jusqu'à devenir de septembre 2011 à octobre 2014, le Directeur du département Informatique de Polytech Nantes. La spécialité Informatique de Polytech Nantes compte près de 210 étudiants (3 promotions de 70 étudiants). A la formation initiale ingénieur offrant 3 options, s'adossaient un diplôme de formation continue (jusqu'en 2018) et 3 Masters (jusqu'en 2017). L'équipe pédagogique compte 21 enseignants-chercheurs, 1 PAST, 2 secrétaires et plus de 30 vacataires. A noter : ce mandat à la direction du département informatique de Polytech Nantes a été marqué par la demande de réhabilitation de la spécialité faite auprès de la CTI (Commission des Titres d'Ingénieur), avec pendant la période 2012-2013 la construction du dossier, puis sa défense en janvier 2014. Finalement la spécialité a obtenu sa réhabilitation pour 6 ans.

Depuis 2017 je suis co-responsable du parcours international VICO « Visual Computing » du master en informatique de l'université de Nantes. Je suis également responsable du parcours « D INFO » du cycle préparatoire PEIP, ce parcours vise les bacheliers STI2D voulant rejoindre Polytech, il est organisé en lien avec le département informatique de l'IUT de Nantes.

Synthèse de mes responsabilités organisationnelles liées à la pédagogie :

— A l'université de Nantes

**Depuis 2017** Co-responsable du parcours VICO master en informatique, et responsable du parcours PEIP-D INFO.

**2011 - 2014** Direction du département Informatique de Polytech Nantes

**2004 - 2009** Chargé de mission (organisation du Comité des utilisateurs informatique)

**2002 - 2014** Membre de Comités de sélection pour des recrutements (Maîtres de conférences, ingénieur d'étude, PAST)

— A l'université de Toulon et du Var

**2000 - 2002** Direction du département SRC de l'IUT

**2000 - 2002** Membre de Commissions mixtes pour des recrutements (Maitres de conférences, PRAG, PAST)

**1997 - 2000** Direction des études

**1997 - 1999** Responsable du suivi des stages



1998 - 2002 Élu au Conseil de l'IUT

## 1.5 Activités en Recherche

Mes activités de recherche au SIS à Toulon ne se sont pas poursuivies dans le domaine du codage vidéo, spécialité de mon doctorat. J'ai alors travaillé dans le cadre de partenariats, d'abord avec le Centre Hospitalier Intercommunal (CHI) de Fréjus-St Raphaël sur le traitement et la communication d'images médicales, puis avec l'Ifremer (centre de Toulon - la Seyne sur Mer) sur des problématiques de trajectographie.

Depuis que j'ai rejoint après mutation en 2002 l'université de Nantes, j'ai retrouvé le domaine du codage des images et de la vidéo (2D ou 3D). Les travaux dans l'équipe IPI du LS2N (nouvelle mouture de l'équipe IVC de l'IRCCyN) se caractérisent par la prise en compte de l'observateur pour optimiser la qualité des images perçues. Dans ce cadre, mes activités scientifiques sont à l'intersection de plusieurs disciplines et plus particulièrement liées à l'informatique par le développement logiciel d'outils et de méthodes pour le codage vidéo, les approches intègrent des modèles de la perception visuelle afin, par exemple, de piloter des algorithmes de codage avancé, ou bien d'optimiser le rendu visuel de scènes 2D ou 3D. J'ai donc demandé à rejoindre en 2005 la 27<sup>e</sup> section car mon ancrage dans la section informatique s'était fait rapidement au niveau de mes activités pédagogiques et plus progressivement au niveau de mes activités de recherche. Mes activités de recherche se sont structurées avec ma participation à de nombreux projets scientifiques, avec successivement depuis 2005, deux projets ANR et un projet européen ITN dans le domaine du codage vidéo, et (actuellement) un projet ANR dans le domaine de l'attention visuelle. J'ai notamment eu la responsabilité scientifique du projet ANR blanc «PERSEE » qui a duré de 2009 à 2013.

Synthèse de mes thématiques de recherche :

- A l'Université de Nantes (depuis 2002) à l'IRCCyN puis au LS2N :
  - Quantification vectorielle pour l'analyse, la représentation efficace et la compression de contenus 3D (nuages de points) [37, 33, 35]
  - Fidélité perceptuelle et codage de textures dynamiques [4, 8, 9, 10, 11, 12, 49, 13, 32, 14]
  - Modèle computationnel de l'attention visuelle [1, 5, 2, 15, 16, 3, 34, 17, 18, 38]
  - Analyse d'images par ondelettes orientées [39, 22, 41, 23, 42]
  - Pré-analyse de la vidéo et stratégie de codage avancées [19, 40, 20, 21, 24]
  - Transformation Mojette et codage source [31, 7, 36]
- A l'Université de Toulon au SIS (1997-2002) :
  - Trajectographie (fusion multi-trajectoires, trajectographie sur base ultra-courte) [25]
  - Détection et suivi d'objets déformables au sein d'images médicales (contours actifs) [43, 26, 45]

- Communication d'images médicales sur réseau [44]

### 1.5.1 Encadrements de thèses et de masters

#### Encadrement doctoral

5 thèses soutenues, une en cours.

**Amira Filali** « *Représentation efficace des contenus dynamiques discrets pour la réalité mixte* »

- Durée : depuis janvier 2018 ; Taux : 60%
- Sujet : Nous nous intéressons à la représentation efficace des contenus immersifs (nuages de points), pour la réalité augmentée et celle mixte, afin de mieux les comprimer.
- Contexte : Contrat doctoral (allocation ministérielle)
- Production commune : 1 conférence nationale [37]

**Karam Naser** « *Fidélité perceptuelle et codage de textures dynamiques* »

- Durée : 3/2014 - 5/2017 ; Taux : 50%
- Sujet : Nous nous intéressons à améliorer la fidélité perceptuelle dans le rendu de textures dynamiques décodées, le codage vidéo est effectué via HEVC.
- Contexte : Projet européen ITN PROVISION
- Production commune : 7 conférences internationales [8, 9, 10, 11, 12, 13, 14], 1 conférence nationale [32], 1 chapitre de livre [4], et 1 brevet [49]

**Junle Wang** « *From 2D to Stereoscopic-3D Visual Saliency : Revisiting Psychophysical Methods and Computational Modeling* »

- Durée : 10/2009 - 11/2012 ; Taux : 50%
- Sujet : Dans le domaine de l'attention visuelle en 3D, la thèse a comparé les vérités de terrain, conçu et exploité des tests oculométriques binoculaires, et proposé un modèle computationnel de l'attention visuelle 3D.
- Contexte : Projet ANR PERSEE
- Production commune : 2 articles en revue [2, 3], 1 chapitre de livre [5], 4 conférences internationales [15, 34, 17, 18], 4 rapports ANR [130, 206, 207, 240]

**Olivier Brouard** « *Pré-analyse de la vidéo pour un codage adapté. Application au codage de la TVHD en flux H.264* »

- Durée : 03/2006 – 07/2010 ; Taux : 70%
- Sujet : Cette thèse a proposé une méthode de pré-analyse de la vidéo avant son encodage en flux H.264, la méthode de pré-analyse intégrait un modèle d'attention visuelle.
- Contexte : Projet ANR ArchiPEG

- Production commune : 3 conférences internationales [19, 20, 21], 2 conférences nationales [38, 40], 4 rapports ANR [274, 275, 312, 339]

**Guillaume Jeannic** « *Représentation structurelle d’images par transformées locales en ondelettes orientées et codage* »

- Durée : 10/2004 – 11/2008 ; Taux : 70%
- Sujet : Dans un contexte de compression d’images, cette thèse a étudié les transformations en ondelettes directionnelles implantées via schéma lifting. Un schéma de filtrage orienté suivant les directions locales régulières a été proposé.
- Production commune : 2 conférences internationales [22, 23], 3 conférences nationales [39, 41, 42]

**Pierre Verbert** « *Sur la redondance des transformations Mojette en dimension  $n$  et en ligne* »

- Durée : 10/1999 – 9/2004 ; Taux : 20%
- Sujet : Cette thèse a exploré l’extension de la transformation Mojette à des dimensions supérieures à 2 et à des projections incomplètes.
- Production commune : 1 conférence internationale [36]

### Encadrement de masters de recherche

9 Masters encadrés et 7 co-encadrés.

**2019** Tianjian Feng : « *Tree-Structured Point-Lattice Vector Quantization and Graph Transform for Multi-scale Representation of 3D Point Cloud* » ; Taux : 100%

**2017** Zuxu Shi : « *Quantification vectorielle algébrique et arborescente de nuages de points pour la réalité virtuelle* » ; Taux : 100%

**2016** Chengyue Ma : « *Optimisation débit-distorsion avec critère perceptuel, application à HEVC* » ; Taux : 100%

**2014** Lei Xu : « *Conception d’une métrique (basée MFS) pour évaluer la qualité de textures dynamiques décodées* » ; Taux : 50%

**2013** Sen Liu : « *Transformée Mojette avec des réseaux cubiques non classiques* » ; Taux : 50%

**2010** Luo Siya : « *Visualisation 3D de molécules par quantification vectorielle avec « lattices » (réseaux réguliers de points)* » ; Taux : 100%

**2009** Junle Wang : « *Développement d’un nouveau modèle d’attention visuelle* » ; Taux : 30%

**2008** Xiaole Fang : « *Calcul de cartes d’importance visuelle de vidéos HD pour leur encodage en flux H.264/AVC* » ; Taux : 100%

**2007** Jérôme Gorin : « *Pré-analyse de plans vidéo HD. Application à l’optimisation du codage en flux H.264* » ; Taux : 100%

- 2007** Matthieu Moinard : « *Quantification et codage après transformées en ondelettes orientées : Application au codage d'images fixes* » ; Taux : 100%
- 2007** Yin Xu : « *Crypto-compression basée chaos des images en vue de la transmission* » ; Taux : 30%
- 2005** Sylvain Tourancheau : « *Prise en compte de la géométrie de l'image avant analyse par ondelettes dans le codage vidéo* » ; Taux : 100%
- 2004** Batista Casanelles : « *Indexation de mosaïques vidéo 1D par Quantification Vectorielle algébrique et arborescente* » ; Taux : 100%
- 2004** Stéphane Péchard : « *Utilisation de modèles psychovisuels pour le contrôle de la prédiction spatiale de la norme de compression vidéo H.264/AVC* » ; Taux : 50%
- 2000** Jean Michelis : « *Développement d'outils d'aide au télédiagnostic dans un réseau d'images médicales* » ; Taux : 50%
- 1999** Steeve Nicolay : « *Reconstruction 3D de l'aorte issue d'une séquence d'images tomodensitométriques par contours actifs* » ; Taux : 50%

### Encadrement de mémoires CNAM

4 Mémoires CNAM encadrés (dont 3 en lien avec des problématiques de recherche).

- 2011** Cédric Ramassamy : « *Quantification Vectorielle Algébrique et Arborescente pour l'analyse de données moléculaires* » ; Taux : 100%
- 2011** Xavier Pineau : « *Conception d'outils avancés pour le codage de la vidéo 3D : Fusion d'images couleur binoculaires perçues* » ; Taux : 100%
- 2007** Maurice Ngom : « *Etude, conception et réalisation d'une architecture réseau complète à base de logiciels libres* » ; Taux : 100%
- 2006** Bertrand Ouary : « *Etude d'un schéma de pré-analyse de la vidéo afin d'optimiser sa compression avec le codec H.264/AVC* » ; Taux : 100%

### 1.5.2 Autres activités scientifiques

Depuis 2002 ma recherche s'est structurée autour de nombreux projets scientifiques que la première partie (I) de ce mémoire détaille. Je donne donc simplement ci-après quelques données synthétiques sur le projet ANR blanc PERSEE que j'ai eu la responsabilité de coordonner, et sur les autres projets auxquels j'ai contribué.

J'ai aussi participé à de nombreux autres aspects de la vie d'un chercheur par des invitations (relectures d'articles, participation à des comités de programme, participation à des comités de sélection pour des recrutements, expertises nationales, régionales et industrielles, jurys de thèses, membre de Conseils de laboratoire ou de la recherche, collaborations scientifiques internationales, actions de vulgarisation) et des animations (organisation de colloques, de réunions de GDR, et de séminaires),

## Responsabilités scientifiques

### Coordination du projet ANR PERSEE « *Schémas perceptuels & codage vidéo 2D et 3D* »

- Sujet : Modélisation et représentation pour un codage basé qualité perceptuelle de la vidéo naturelle 3D
- Type de projet : ANR Programme blanc, projet PERSEE (réf. ANR-09-BLAN-0170)
- Période : 2009-2013
- Partenaires : IRCCyN, IETR-Rennes, INRIA-Rennes, LTCI-TelecomParisTech
- Résultats : 4 thèses soutenues, 10 publications en revues, 3 brevets

## Participations à des projets

### Participation au projet ANR DISSOCIE « *Détection automatique des Saillances du point de vue Opérateurs et Compression Intelligente des vidéos de drones* »

- Sujet : Méthodes de détection d'objets saillants du point de vue des opérateurs dans des vidéos capturées par des drones, puis compression intelligente de ces vidéos
- Type de projet : ANR Programme ASTRID, projet DISSOCIE (réf. ANR-17-ASTR-0009)
- Période : 2018-2021 (en cours)
- Partenaires : IETR-Rennes, INRIA-Rennes, LS2N

### Participation au projet ITN PROVISION « *PeRceptually Optimised Video Compression* »

- Sujet : Compression vidéo optimisée perceptuellement
- Type de projet : FP7 Programme européen ITN (Marie Skłodowska-Curie Initial Training Networks, appel FP7-PEOPLE-2013-ITN, n° 608231)
- Période : 2013-2017
- Partenaires : Fraunhofer Heinrich-Hertz-Institut (HHI, Allemagne), Université de Bristol (UK), British Broadcasting Corporation (BBC, UK), Université de Aachen (RWTH, Allemagne), Université de Nantes (IRCCyN)
- Partenaires associés : Queen Mary University of London, Technicolor (France), Vitec (France),

### Participation au projet ANR ArchiPEG « *Architecture modulaire, programmable, évolutive, graduable, pour les applications en vidéo numérique en très hautes performances* »

- Sujet : Conception d'une nouvelle architecture de traitement de la vidéo apte à traiter les algorithmes de compression de la TVHD
- Type de projet : ANR programme RIAM, projet ArchiPEG, (réf. ANR-05-RIAM-01401)

- Période : 2005-2008
- Partenaires : Vitec Multimédia, IRCCyN

**Participation au projet RNTL « *Domus Videum* »**

- Sujet : Sélection, indexation, visualisation avancée et recherche par le contenu des documents vidéo pour les plates-formes multimédia domestiques avec apprentissage du profil utilisateur
- Période : 2002-2004
- Partenaires : Thomson Multimédia, IRCCyN, IRISA, IRIN, SFRS, INA, LaBRI

**Participation au projet ORACLE « *Outils pour la Réalisation Assistée de Cartographies d'images sous-marines bien Localisées pour l'Environnement* »**

- Sujet : Reconnaissance de l'environnement littoral par des moyens d'investigations optiques ou acoustiques, aériens sub-surface ou sous-marin, et d'établir des données cartographiques de référence
- Période : 2001-2002
- Partenaires : Ifremer, Université de Toulon et du Var, Systemia, ACSA, PNPC, MATRA

**Collaborations avec des chercheurs invités**

**Bogdan Cramariuc** IT Center for Science and Technology (CITST, Roumanie)

- Période : un mois en 2010 et un autre en 2011
- Sujet : « Analyse de données biomoléculaires après quantification vectorielle »

**James Fowler** Department of Electrical and Computer Engineering, Mississippi State University

- Période : un mois en 2009
- Sujet : lié au « Compress sensing »

**Expertises et collaborations avec l'environnement socio-économique**

**Expertise nationale** pour l'ANR : projet déposé en 2011, et expertise CIFRE en 2017

**Expertise régionale** pour l'Ifremer (Laboratoire de Robotique sous-marine, La Seyne sur Mer)

- Durée : 4 mois en 2000
- Sujet : expertise pour le système de trajectographie sur base ultra-courte
- Participants : Vincent Ricordel (50%), Claude Jauffret (50%)

**Expertise industrielle** pour la société VITEC Multimédia (Châtillon)

- Période : 4 mois en 2005
- Sujet : Expertise du codeur MPEG-2 de Vitec Multimédia
- Participants : Vincent Ricordel (50%), Dominique Barba (50%)

**Expertise industrielle** pour la société Courant Multimédia (CMM à Poitiers)

- Période : 2 mois en 2004
- Sujet : Expertise pour l'étude de faisabilité d'un module de transmission multimédia sur courant porteur
- Participants : Vincent Ricordel (25%), Nicolas Normand (50%), Patrick Le Callet (25%)

**Membre de Conseils et de Comités**

**Membre nommé du Conseil du laboratoire IRCCyN** de 2002 à 2011

**Membre nommé du Conseil de la recherche de Polytech Nantes** de 2002 à 2007

**Membre élu de la Commission de spécialistes 61/63** de 2000 à 2002 et à l'université de Toulon

**Participations à des Comités de sélection** pour le recrutement de maîtres de conférence à Toulon (2), à Nantes (5) et à Rennes (2)

**Animations de la recherche**

**Organisation de séminaires :**

**Co-organisation de journées du GDR ISIS**

8 juillet 2016 - thème : « Optimisation (perceptuelle) d'HEVC et des futurs standards de compression vidéo » à Polytech Nantes

19 mars 2019 - thème : « Compression et qualité des contenus 360, Light Field et Point Cloud (3D) » à l'INSA Rennes

**Organisation des séminaires de l'équipe IPI (2017-2019)**

**Organisation de colloques :**

**Co-organisation d'une session spéciale à PCS** (« Picture Coding Symposium »), titre de la session « Challenges in Perceptual Compression of Video » à Nuremberg (5 décembre 2016)

**Présidence des sessions poster du SIFWICT** (« Sino-French Workshop on Information and Communication Technologies ») à Nantes (12 juin 2015)

**Gestion pour le Workshop « Packet Video »** à Nantes (28-29 avril 2003)

**Secrétariat du « HCM MAVI Course Workshop »** (HCM : EU Human Capital and Mobility Program; MAVI : Model based Analysis of Video Information) à Pori en Finlande (17-18 avril 1997)

**Participations à des comités de programme :**

**SIFWICT** en 2015

**Conférences CORESA** « Compression et représentation des signaux audiovisuels »  
(2007, 2010, 2013 et 2018)

**IPTA** « International Conference on Image Processing Theory, Tools and Applications  
» (2010 et 2012)

**Workshop « Packet Video »** en 2003

**Examinations pour des jurys de thèses :**

**KhouLOUD Samrouth** (INSA de Rennes, 2014) Titre : « Représentation et Compression à haut niveau sémantique d'images 3D »

**Benjamin Le Guen** (Université de Rennes, 2008) Titre : « Adaptation du contenu spatio-temporel des images pour un codage par ondelettes »

**Olivier Crave** (Ecole Nationale des Télécommunications, 2008) Titre : « Approches théoriques en codage vidéo robuste multi-terminal »

**Relectures pour des revues :**

**TCSVT** « IEEE Transactions on Circuits and Systems for Video Technology » (5)

**TSP** « IEEE Transactions on Signal Processing » (1)

**TIP** « IEEE Transactions on Image Processing » (1)

**TM** « IEEE Transactions on Multimedia » (1)

**IRBM** « Elsevier - Innovation and Research in BioMedical Engineering » (2)

**Relectures pour des conférences nationales et internationales :**

**ICIP** « IEEE International Conference on Image Processing » (52)

**PCS** « Picture Coding Symposium » (10)

**GRETSI** « Colloque GRETSI de traitement du signal et des images » (5)

**IPTA** « International Conference on Image Processing Theory, Tools and Applications  
» (5)

**CORESA** Conférence « Compression et représentation des signaux audiovisuels » (10)

**SIFWICT** Conférence « Sino-French Workshop on Research Collaborations in Informations and Communication Technologies » (5)

**Membre de Comités de suivi de thèses :**

**Kevin Reuze** Titre : « Towards Optimal Predictive Coding of Intra Prediction Modes in the Future Video Coding » (2017-2019)

**Saurabh Puri** Titre : « Optimisation des transformées pour la compression vidéo » (2015, 2016)



**Filippo Mazza** Titre : « Using social networks for face pictures aesthetic estimation and interpretation » (2012-2015)

**Jing LI** Titre : « Characterization of Depth Perception and Visual Comfort for Controlling The QoE on 3D-TV » (2007-2010)

**Participation régulière à des événements de vulgarisation, par exemple :**

**Portes ouvertes de Polytech Nantes**

**Fête de la science** (2015, 2016)

**Nuit de la recherche européenne** (2015, 2016)

### 1.5.3 Publications

Ma liste des publications comprend :

3 revues internationales à comité de lecture

3 chapitres d'ouvrages (1 en français et 2 en anglais)

24 conférences internationales à comité de lecture et actes publiées

6 workshops internationaux

12 conférences nationales à comité de lecture et actes publiées

1 brevet

Plus de 20 rapports liés aux projets

1 autre (article de vulgarisation pour une revue)

#### Revues internationales à comité de lecture

- [1] Vassilios KRASSANAKIS, Matthieu Perreira Da SILVA et Vincent RICORDEL. « Monitoring Human Visual Behavior during the Observation of Unmanned Aerial Vehicles (UAVs) Videos ». In : *Drones* 2.36 (2018).
- [2] Junle WANG, Matthieu Perreira Da SILVA, Patrick Le CALLET et Vincent RICORDEL. « A computational model of stereoscopic 3D visual saliency ». In : *IEEE Transaction on Image Processing* 22.6 (juin 2013), p. 2151–2165.
- [3] Junle WANG, Patrick Le CALLET, Sylvain TOURANCHEAU, Vincent RICORDEL et Matthieu Perreira Da SILVA. « Study of depth bias of observers in free viewing of still stereoscopic synthetic stimuli ». In : *Journal of Eye Movement Research* 5.5 (sept. 2012), p. 1–11.

**Chapitres d'ouvrages collectifs**

- [4] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Perceptual Texture Similarity for Machine Intelligence Applications ». In : *Visual Content Indexing and Retrieval with Psycho-Visual Models*. Sous la dir. de Jenny BENOIS-PINEAU et Patrick Le CALLET. Springer, 2017.
- [5] Vincent RICORDEL, Junle WANG, Matthieu Perreira Da SILVA et Patrick Le CALLET. « 2D & 3D Visual Attention for Computer Vision : Concepts, Measurement, and Modeling ». In : *Innovative Research in Attention Modeling and Computer Vision Applications*. Sous la dir. de Rajarshi PAL. IGI Global publication, 2015. Chap. 1.
- [6] Marc ANTONINI et Vincent RICORDEL. « Quantification ». In : *Compression et codage des images et des vidéos*. Sous la dir. de Michel BARLAUD et Claude LABIT. Traité IC2, Série traitement du signal et de l'image ISBN : 2-7462-0328-6. Hermes Science, jan. 2002. Chap. 2.

**Conférences internationales à comité de lecture et actes publiées**

- [7] Vincent RICORDEL, Nicolas NORMAND et Jeanpierre GUÉDON. « Mojette Transform on Densest Lattices in 2D and 3D ». In : *International Conference on Discrete Geometry for Computer Imagery (DGCI)*. Sous la dir. de SPRINGER. Lecture Notes in Computer Science 10502. Vienne, Austria. Sept. 2017, p. 159–170.
- [8] Chengyue MA, Karam NASER, Vincent RICORDEL, Patrick Le CALLET et Chunmei QING. « An adaptive Lagrange multiplier determination method for dynamic texture in HEVC ». In : *IEEE International Conference on Consumer Electronics-China (ICCE-China)*. Guangzhou, China. Déc. 2016, p. 1–4.
- [9] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « A foveated short term distortion model for perceptually optimized dynamic textures compression in HEVC ». In : *Picture Coding Symposium (PCS)*. Nuremberg, Germany. Déc. 2016.
- [10] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Estimation of Perceptual Redundancies of HEVC Encoded Dynamic Textures ». In : *International Conference on Quality of Multimedia Experience (QoMEX)*. Lisbon, Portugal. Juin 2016.
- [11] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Modeling the Perceptual Distortion of Dynamic Textures and Its Application in HEVC ». In : *IEEE International Conference on Image Processing (ICIP)*. Phoenix, Arizona, United States. Sept. 2016.
- [12] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Local Texture Synthesis : A Static Texture Coding Algorithm Fully Compatible With HEVC ». In : *International Conference on Systems Signals and Image Processing (IWSSIP)*. London, UK. Sept. 2015.

- [13] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Performance Analysis of Texture Similarity Metrics in HEVC Intra Prediction ». In : *Video Processing and Quality Metrics for Consumer Electronics (VPQM)*. Chandler, Arizona, USA. Fév. 2015.
- [14] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Experimenting Texture Similarity Metric STSIM for Intra Prediction Mode Selection and Block Partitioning in HEVC ». In : *Digital Signal Processing (DSP)*. Hong Kong, Chine. Août 2014.
- [15] Junle WANG, Matthieu Perreira Da SILVA, Patrick Le CALLET et Vincent RICORDEL. « Study of center-bias in the viewing of stereoscopic image and a framework for extending 2D visual attention models to 3D ». In : *SPIE Electronic Imaging*. San Francisco, USA. Fév. 2013, p. 3–7.
- [16] Matthieu URVOY, Marcus BARKOWSKY, Romain COUSSEAU, Yao KOUDOTA, Vincent RICORDEL, Patrick Le Callet and Jesús GUTIÉRREZ et Narciso GARCIA. « NAMA3DS1-COSPAD1 : Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences ». In : *International Workshop on Quality of Multimedia Experience (QoMEX)*. Yarra Valley, Australie. Juil. 2012.
- [17] Junle WANG, Marcus BARKOWSKY, Vincent RICORDEL et Patrick Le CALLET. « Quantifying how the combination of blur and disparity affects the perceived depth ». In : *SPIE Electronic Imaging*. San Francisco Airport, California, USA. Jan. 2011.
- [18] Junle WANG, Patrick Le CALLET, Vincent RICORDEL et Sylvain TOURANCHEAU. « Quantifying depth bias in free viewing of still stereoscopic synthetic stimuli ». In : *European Conference on Eye Movements (ECEM)*. Marseille, France. Août 2011.
- [19] Olivier BROUARD, Fabrice DELANNAY, Vincent RICORDEL et Dominique BARBA. « Spatio-Temporal Segmentation and Regions Tracking of High Definition Video Sequences using a Markov Random Field Model ». In : *IEEE International Conference on Image Processing (ICIP)*. San Diego, Californie, USA. Oct. 2008.
- [20] Olivier BROUARD, Fabrice DELANNAY, Vincent RICORDEL et Dominique BARBA. « Fast Long-Term Motion Estimation for High Definition Video Sequences Based on Spatio-Temporal Tubes and using the Nelder-Mead Simplex Algorithm ». In : *Picture Coding Symposium (PCS)*. Lisbonne, Portugal. Nov. 2007.
- [21] Olivier BROUARD, Fabrice DELANNAY, Vincent RICORDEL et Dominique BARBA. « Robust Motion Segmentation for High Definition Video Sequences using a Fast Multi-Resolution Motion Estimation based on Spatio-Temporal Tubes ». In : *Picture Coding Symposium (PCS)*. Lisbonne, Portugal. Nov. 2007.
- [22] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « A Multiresolution Approach for the Coding of Edges of Still Images using Adaptive Arithmetic Coding ». In : *Picture Coding Symposium (PCS)*. Lisbonne, Portugal. Nov. 2007.

- [23] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « The Edge Driven Oriented Wavelet Transform : An Anisotropic Multidirectional Representation With Oriented Lifting Scheme ». In : *Visual Communications and Image Processing (VCIP) part of the SPIE Symposium on Electronic Imaging*. San Jose, Californie, USA. Jan. 2007.
- [24] Vincent RICORDEL, Patrick Le CALLET, Mathieu CARNEC et Dominique BARBA. « Flexible Storage of Still Images with a Perceptual Quality Criterion ». In : *Advanced Concepts for Intelligent Vision Systems (ACIVS)*. Anvers, Belgique. Sept. 2005.
- [25] Vincent RICORDEL, Sébastien PARIS et Jan OPDERBECKE. « Trajectory Estimation for Ultrashort Baseline Acoustic Positioning Systems ». In : *IEEE Aerospace Conference*. Big Sky, Montana, USA. Mar. 2001.
- [26] William PUECH, Guy PASSAIL et Vincent RICORDEL. « Analysis and Optimisation of 3D Reconstruction Method of the Aorta from a Tomographic Image Sequence ». In : *European Signal Processing Conference (EUSIPCO)*. Tampere, Finlande. Sept. 2000.
- [27] Vincent RICORDEL et Moncef GABBOUJ. « Adaptive Tree-Structure Lattice Vector Quantization for Video Coding ». In : *European Signal Processing Conference (EUSIPCO)*. Rhodes, Grèce. Sept. 1998.
- [28] Vincent RICORDEL et Claude LABIT. « Tree-Structured Lattice Vector Quantization ». In : *European Signal Processing Conference (EUSIPCO)*. Trieste, Italie. Sept. 1996.
- [29] Vincent RICORDEL et Claude LABIT. « Vector Quantization by Hierarchical Packing of Embedded Truncated Lattices ». In : *Visual Communications and Image Processing (VCIP)*. Taiwan, Chine. Mai 1995.
- [30] Vincent RICORDEL et Claude LABIT. « Vector Quantization by packing of embedded truncated lattices ». In : *IEEE International Conference on Image Processing (ICIP)*. Washington DC, USA. Oct. 1995.

### Workshops internationaux

- [31] Vincent RICORDEL, Nicolas NORMAND et Jeanpierre GUÉDON. « Analysis of Mottlet Transform Projections using Z3 lattice versus A3 lattice ». In : *Sino-French Workshop on Research Collaborations in Informations and Communication Technologies (SIFWICT)*. Qingdao, China. Juin 2017.
- [32] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Texture Similarity Metrics Applied to HEVC Intra Prediction ». In : *Sino-French Workshop on Research Collaborations in Informations and Communication Technologies (SIFWICT)*. Nantes, France. Juin 2015.

- [33] Cédric RAMASSAMY, Vincent RICORDEL, Oana CRAMARIUC et Bogdan CRAMARIUC. « Lattice Vector Quantization for the Analysis of Molecular Data ». In : *Signal Processing and Applied Mathematics for Electronics and Communications (SPAMEC)*. Cluj-Napoca, Roumanie. Août 2011.
- [34] Junle WANG, Marcus BARKOWSKY, Vincent RICORDEL et Patrick Le CALLET. « Quantifying how the combination of blur and disparity affects the perceived depth ». In : *Sino-French Workshop on Research Collaborations in Informations and Communication Technologies (SIFWICT)*. Nantes, France. Mai 2011.
- [35] Vincent RICORDEL, Florent AUTRUSSEAU, William DUPUY et Dominique BARBA. « 1D-Mosaics Grouping using Lattice Vector Quantization for a Video Browsing Application ». In : *International Workshop on Content-Based Multimedia Indexing (CBMI)*. Riga, Lettonie. Juin 2005.
- [36] Pierre VERBERT, Vincent RICORDEL et Jeanpierre GUÉDON. « Analysis of Mojette Transform Projections for an efficient Coding ». In : *Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*. Lisbonne, Portugal. Avr. 2004.

#### Conférences nationales à comité de lecture et actes publiées

- [37] Amira FILALI, Vincent RICORDEL et Nicolas NORMAND. « Tree-Structured Point-Lattice Vector Quantization for 3-D Point Cloud Geometry Compression ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Nov. 2018.
- [38] Olivier BROUARD, Vincent RICORDEL et Dominique BARBA. « Cartes de Saillance Spatio-Temporelle basées Contrastes de Couleur et Mouvement Relatif ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Toulouse, France. Mar. 2009.
- [39] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « Étude comparative de techniques de lifting orientées sur des contenus synthétiques représentant des structures rectilignes ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Toulouse, France. Mar. 2009.
- [40] Olivier BROUARD, Fabrice DELANNAY, Vincent RICORDEL et Dominique BARBA. « Estimation Robuste du Mouvement Global pour des Séquences Vidéo Haute Définition en utilisant une Estimation Multi-Résolutions du Mouvement ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Montpellier, France. Nov. 2007.
- [41] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « Représentation structurelle d'images fixes par transformée en ondelettes orientées ». In : *Colloque GRETSI sur le traitement du signal et des images (GRETSI)*. Troyes, France. Sept. 2007.

- [42] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « Extraction de contours multirésolution pour un codage d’images par bandelettes ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Caen, France. Nov. 2006.
- [43] Jean MICHELIS, Vincent RICORDEL, William PUECH, Guy PASSAIL et Michel DUMAS. « Intégration d’applet JAVA dans un réseau d’images médicales : aide au télédiagnostic ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Poitiers, France. Oct. 2000.
- [44] Guy PASSAIL, Vincent RICORDEL, Alexander KOLESNIKOV et William PUECH. « Mise en œuvre d’un réseau pour la communication d’images médicales ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Nice-Sophia-Antipolis, France. Juin 1999.
- [45] William PUECH, Guy PASSAIL, Steeve NICOLAY et Vincent RICORDEL. « Analyse et amélioration de méthodes de reconstruction 3D de l’aorte à partir d’une séquence d’images tomodensitométriques ». In : *Colloque GRETSI sur le traitement du signal et des images (GRETSI)*. Vannes, France. Sept. 1999.
- [46] Vincent RICORDEL, Moncef GABBOUJ et Claude LABIT. « Quantification vectorielle algébrique et arborescente : méthode d’actualisation du dictionnaire. Application au codage vidéo ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Lannion, France. Juin 1998.
- [47] Vincent RICORDEL et Claude LABIT. « Quantification vectorielle algébrique et arborescente ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Grenoble, France. Fév. 1996.
- [48] Vincent RICORDEL et Claude LABIT. « Quantification vectorielle par emboîtement de réseaux réguliers de points ». In : *Colloque GRETSI sur le traitement du signal et des images (GRETSI)*. Juans-les-Pins, France. Sept. 1995.

### Brevet

- [49] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. *Method for encoding video frames based on local texture synthesis and corresponding device*. France, Patent n° : 15306367.2 - 1908. 2015, European Patent Office (Munich, Germany). 2015.

## Chapitre 2

# Introduction générale

### 2.1 Positionnement du travail

Ce mémoire retrace mon itinéraire scientifique depuis 2002, année de mon arrivée à Polytech au sein de l'équipe de recherche IVC (Images & Vidéo Communications) de l'IRCCyN (Institut de Recherche en Communications et Cybernétique de Nantes, UMR 6597), cette équipe est devenue en 2017 équipe IPI (Image Perception Interaction)<sup>1</sup> au sein du nouveau laboratoire LS2N<sup>2</sup> (Laboratoire des Sciences du Numérique de Nantes, UMR 6004). Ce nouveau poste, obtenu après mutation, marqua mon retour à des activités de recherche dans le domaine du codage des images et de la vidéo, et le contexte particulier de l'équipe IVC fut déterminant en orientant mon parcours vers un codage perceptuel. En effet les travaux de recherche qui y étaient menés s'organisaient autour de thématiques complémentaires et cohérentes, et les deux qui me concernaient plus particulièrement étaient, d'un coté « Perception et Modèles psychovisuels » et de l'autre « Représentation et communication Vidéo-Multimédia ». L'équipe IPI ayant conservé ces 2 thématiques (seule la première a été renommée « Représentation et perception : modèles psycho-visuels »), mes activités de recherche en codage perceptuel s'y sont naturellement prolongées.

Mes travaux se sont donc inscrits dans ces thématiques où, brièvement, l'objectif est de trouver la meilleure représentation possible de l'information vidéo, représentation qu'il sera ensuite possible de comprimer efficacement à des fins de stockage ou de transmission, la finalité étant que les vidéos codées-décodées soient perçues sans défaut par les observateurs (on parle ainsi de fidélité perceptuelle sans perte). Sachant que tous les standards de compression vidéo actuels sont destructifs, c'est-à-dire que pour atteindre des taux de compression significatifs les algorithmes de codage doivent altérer une partie de l'information (altération que l'on mesure classiquement via une métrique objective de type EQM « Erreur Quadratique Moyenne » ou PSNR « Peak Signal-to-Noise Ratio »).

---

1. Site web de l'équipe IPI : [www.ls2n.fr/equipe/ipi](http://www.ls2n.fr/equipe/ipi)

2. Site web du LS2N : [www.ls2n.fr](http://www.ls2n.fr)

On comprend aisément comment des résultats de recherche fondamentale en modélisation du système visuel humain (SVH) ou liés à la connaissance de mécanismes de la perception visuelle, peuvent irriguer une recherche appliquée en compression vidéo. Car ils contribuent à mettre au point une représentation de l'information basée sur le contenu perçu qui puisse servir de paradigme pour un codage basé qualité perceptuelle des contenus vidéo 2D et 3D. La recherche menée dans les équipes IVC puis IPI, et relative à l'évaluation de la qualité visuelle, contribue aussi à offrir un contexte particulièrement propice au codage vidéo. Les méthodologies de tests et leur environnement (plateformes expérimentales dédiées), permettent en effet d'évaluer la qualité perçue (subjective) des vidéos traitées (décodées), de définir les conditions pour utiliser de nouveaux dispositifs de visualisation, ou pour déployer et piloter de nouveaux services.

Ce mémoire s'intéresse à la recherche en compression vidéo et à son évolution depuis ces 15 dernières années. On ne peut que constater combien la technologie vidéo est devenue pendant cette période omniprésente. Pour le montrer, on peut énumérer (sans être exhaustif) les nouvelles technologies et services qui ont émergé :

- vers 2005 : vidéo mobile, TVHD (terrestre, satellite), Blu-ray, caméras numériques, visiophonie, messagerie multimédia, vidéoconférence, vidéosurveillance ;
- vers 2010 : vidéo 3D, 3DTV, affichage panoramique, vidéo omnidirectionnelle ou FVV « Free Viewpoint Video » ;
- vers 2015 : formats très haute définition UHD « Ultra HD » (4K, 8K), vidéo sur mobile, VoD « Video on Demand » ;
- Actuellement : formats immersifs, vidéo augmentée, virtuelle ou mixte.

Si les besoins des nouveaux services vidéo ont été en partie satisfaits par l'efficacité accrue des réseaux et de leurs technologies de couche physique, en parallèle les codecs vidéo ont été primordiaux en offrant compression et nouvelles fonctionnalités. Toutes ces technologies et services ont en fait réclamé pour leur déploiement, des représentations efficaces des signaux vidéo, et ce sont les standards de compressions successifs qui ont accompagné et permis leur émergence. La recherche en codage vidéo se caractérise alors par un fort contexte normatif où exactement le point de comptabilité se fait au niveau du décodage (c'est la syntaxe du flux comprimé qui est standardisée), la compétition demeurant ouverte au niveau de la conception des procédés d'encodage.

Deux standards aboutis de codage vidéo de la JVT « Joint Video Team »<sup>3</sup> ont marqué la période<sup>4</sup> : d'abord H.264/AVC<sup>5</sup> « Advanced Video Coding » (1<sup>re</sup> édition en 2003, 2<sup>de</sup>

---

3. La JVT est la réunion de l'IUT-T (son groupe VCEG « Video Coding Experts Group ») et de l'ISO/IEC (son groupe MPEG « Moving Picture Experts Group »).

4. Le lecteur intéressé par l'historique des standards et leurs spécifications en trouvera une présentation dans cet ouvrage [93].

5. Standard « ITU-T Rec. H.264 | ISO/IEC 14496-10 » nommé « H.264/AVC » ou « MPEG-4-AVC » ou « MPEG-4 Part 10 », et développé par l'équipe JCT-VC « Joint Collaborative Team on Video Coding » de la JVT.



en 2005) visant la compression efficace des formats vidéo HD « High Definition », il y a eu aussi ses extensions aux formats graduables (SVC « Scalable Video Coding », en 2007) et à ceux multi-vues (MVC « Multiview video coding », en 2009) ; l'autre standard étant HEVC<sup>6</sup> (« High Efficiency Video Coding », en 2013) introduit pour traiter les formats UHD « Ultra HD ». Il y a eu également ses extensions aux formats graduables (SHVC « Scalability extension of HEVC », fin 2014), à ceux multi-vues (MV-HEVC « Multiview extension of HEVC », fin 2014), et à la 3D (3D-HEVC « 3D extension of HEVC », en 2015). Autre nouveauté, depuis 2003 les standards sont accompagnés de leur modèle de vérification (le « JM » puis le « KTA »<sup>7</sup> pour H.264/AVC, le « HM »<sup>8</sup> pour HEVC, et le « HTM » pour ses extensions), permettant aux chercheurs de tester et comparer leurs solutions au sein de logiciels de référence.

L'actuel développement du standard MPEG-I « Coded representation for Immersive Media » pour la compression des contenus immersifs, et en particulier du MPEG-I PCC « Point Cloud Compression » retient toute notre attention. Les travaux sont menés par le JVET « Joint Video Expert Team » (nouveau nom du JVT, équipe conjointe entre les groupes MPEG et VCEG), dans le contexte plus général du développement du futur format de codage vidéo VVC « Versatile Video Coding » qui doit succéder en 2020 à HEVC.

Les performances obtenues avec chaque nouveau standard sont remarquables, typiquement le nouveau codec par rapport au précédent, code avec la même qualité objective (PSNR) un format vidéo donné mais en divisant par deux le débit nécessaire [116, 160, 428]. Ces performances sont dues à des avancées technologiques tirant parties des capacités toujours accrues des calculateurs. Les procédés novateurs sont introduits sur deux plans : soit au niveau de la couche de codage vidéo (qui, on le verra, concerne plus nos travaux), soit au niveau de la syntaxe et de l'architecture de description du flux comprimé. Par exemple H.264/AVC [428] a introduit une combinatoire plus large pour les procédés de prédiction intra-image (plus de directions, un filtre anti-blocs dans la boucle), et inter-images (taille variable des blocs, précision de la compensation du mouvement, références temporelles multiples... ) ; plus de transformations ; des algorithmes de codage entropique plus efficaces ; et une syntaxe de description ayant un niveau d'abstraction élevé par rapport aux couches basses, afin d'assurer flexibilité d'usage et robustesse (face aux erreurs et aux pertes) en transmission. HEVC [164] contient une série d'améliorations incrémentales, on peut lister un partitionnement flexible des blocs basée sur une structure quadtree (CTU) ; plus de modes de prédiction intra ; une plus grande flexibilité dans la taille des blocs transformés ; des approches plus sophistiquées avec interpolation pour la compensation de mouvement ; plus de filtres anti-blocs ; et une syntaxe de description introduisant de nouvelles fonctionnalités et structures facilitant notamment les traitements sur architecture de calculs en parallèle.

Il est cependant frappant de constater que le paradigme de la couche de codage n'a

---

6. Standard « ITU-T H.265 | ISO/IEC 23008-2 » nommé « HEVC » ou « H.265/HEVC », et développé par la JCT-VC.

7. Le JM et le KTA sont disponibles là : [iphome.hhi.de/suehring/tml](http://iphome.hhi.de/suehring/tml)

8. Le HM et HTM sont disponibles là : [hevc.hhi.fraunhofer.de](http://hevc.hhi.fraunhofer.de)

pas changé depuis le standard H.261 [127] (dont la 1<sup>re</sup> version fut publiée en 1988!). Avec un codec dissymétrique, où les décisions sont prises à l'encodage via un schéma hybride intégrant les étapes de prédiction à partir de l'information déjà codée-décodée, puis celles de transformation de blocs 2D, quantification et codage entropique de l'erreur de prédiction. La figure 2.1 l'illustre avec le standard HEVC (certes, comme expliqué précédemment, HEVC affiche un panel complet et complexe de prédicteurs, de filtres, etc.). Il est aussi surprenant de constater la persistance de l'implantation dans les codeurs pour la phase d'optimisation débit/qualité d'un calcul de distorsion simple de type EQM, alors que l'on sait la corrélation pauvre entre cette métrique et le jugement humain [264]. Malgré la conception de nombreuses nouvelles métriques de qualité [361] aucune n'a encore supplantée l'EQM dans les codeurs. Finalement la communauté du codage vidéo s'accorde pour considérer que le meilleur moyen pour accéder aux performances réelles d'un codec est de mesurer la qualité subjective perçue des vidéos décodées par des observateurs humains [127] (et le VQEG<sup>9</sup> « Video Quality Experts Group » définit et normalise depuis 1997 des procédures de tests perceptuels).

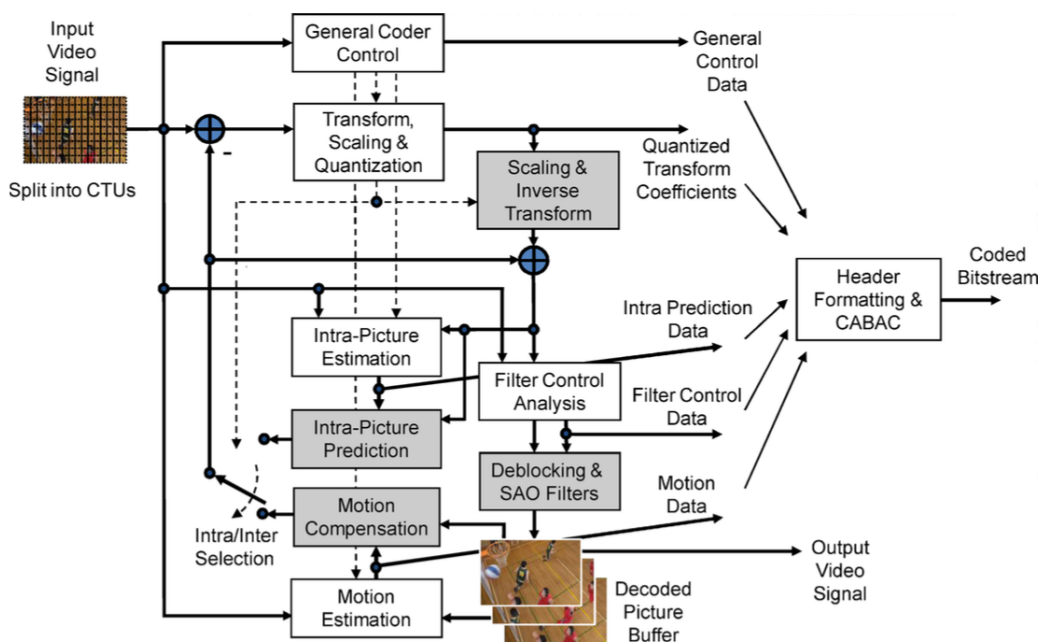


FIGURE 2.1 – Boucle d'encodage hybride type pour HEVC (extrait de [164]).

Pendant la dernière décennie, la recherche en codage vidéo s'est donc mobilisée afin de délivrer des volumes élevés de contenus plus immersifs sur des canaux à bande limitée voire congestionnés. avec l'objectif que les contenus décodés possèdent une qualité subjective irréprochable. Nos travaux se positionnent dans la recherche des outils et des méthodes où les choix, pour améliorer la qualité perceptuelle, guident le codage vidéo.

9. Site web du VQEG : [www.its.bldrdoc.gov/vqeg](http://www.its.bldrdoc.gov/vqeg)

Des approches possibles pour améliorer la couche de codage et/ou offrir de nouvelles fonctionnalités de codage, ont porté sur ses unités « classiques » (transformation, quantification, prédictions intra et inter, métriques de distorsion etc.) tout en tenant compte des contraintes d'un fonctionnement du codeur en boucle fermée. Des approches ont été aussi plus globales visant à optimiser le compromis débit-distorsion du système, d'autres encore ont cherché à ouvrir vers de nouveaux formats et services vidéo (par exemple des modèles d'attention visuel, ou récemment la compression de nuages de points).

## 2.2 Contributions et annonce du plan

Nos travaux se positionnent dans cette recherche d'outils et de méthodes pour améliorer la qualité perceptuelle du codage vidéo 2D et 3D. Nous avons essentiellement travaillé au sein de projets de recherche visant à contribuer aux efforts faits en normalisation. Nous verrons que nous avons agi au niveau de la couche de codage, mais avec différentes approches. Certaines portent sur des unités, qui peuvent être considérées comme « classiques » (la transformation, la quantification, la métrique de distorsion etc.), d'autres moins (les modèles d'attention visuelle, les représentations discrètes 3D etc.), une autre approche a été globale (avec une pré-analyse avant encodage vidéo optimisé). Le plan de mémoire traduit donc cet ensemble de contributions au codage vidéo perceptuel et l'organise en deux parties.

La partie **I** fait une présentation de mes contributions à trois projets de recherche (voir aussi la figure 2.2) : le projet ANR ArchiPEG présenté au chapitre 3, le projet ANR PERSEE vu au chapitre 4, et le projet ITN PROVISION vu au chapitre 5. L'objectif est de montrer le poids important pris par ces projets dans la recherche en codage vidéo dont ils accompagnent et préparent les évolutions<sup>10</sup>. Ainsi ArchiPEG vise l'optimisation du codage du format vidéo HD en flux H.264/AVC, PERSEE qui suit s'intéresse au codage perceptuel des formats vidéo 2D et 3D (donc via le codec H.264 et son extension MVC), enfin PROVISION a pour contexte le codage perceptuel avec HEVC. L'objectif est aussi d'exposer le rôle particulier des équipes IVC puis IPI qui ont apporté un support et une expertise propices au développement des approches perceptuelles. Enfin le but est surtout d'expliquer mes missions et de partager mon expérience, sachant qu'un développement plus important sera fait du projet PERSEE dont j'ai assuré la responsabilité scientifique.

La partie **II** fait une présentation d'une sélection de mes contributions au codage perceptuel menées au travers les thèses et certains masters que j'ai encadrés ou co-encadrés (voir aussi la figure 2.2). Les travaux menés avec Guillaume Jeannic et relatifs à l'étude des transformations orientées sont présentés au chapitre 6. Avec Olivier Brouard (voir le chapitre 7) nous avons travaillé dans le cadre du projet ArchiPEG sur un schéma de pré-analyse avant encodage vidéo. Avec Junle Wang (au chapitre 8) nous avons développé un modèle d'attention visuelle pour la vidéo 3D (stéréo) dans le contexte du

---

10. Je ne présenterai donc pas le projet ANR DISSOCIE (site web : [sites.google.com/insa-rennes.fr/projetanrastrid-dissocie](https://sites.google.com/insa-rennes.fr/projetanrastrid-dissocie)) ayant pour objet la détection automatique de saillances dans des vidéos de drones et qui a débuté en 2018, car je n'y contribue pas sur une tâche liée au codage vidéo.

projet PERSEE. Avec Karam Naser (au chapitre 9), nous avons développé et exploité la notion de similarité entre textures visuelles, le travail a été mené dans le cadre du projet PROVISION. Enfin avec Amira Filali (au chapitre 10), nous travaillons sur une compression de nuages de points 3D basée sur la quantification vectorielle. La conclusion générale de ce mémoire est faite au chapitre 11, des perspectives et mes projets à moyens et longs termes sont alors donnés.

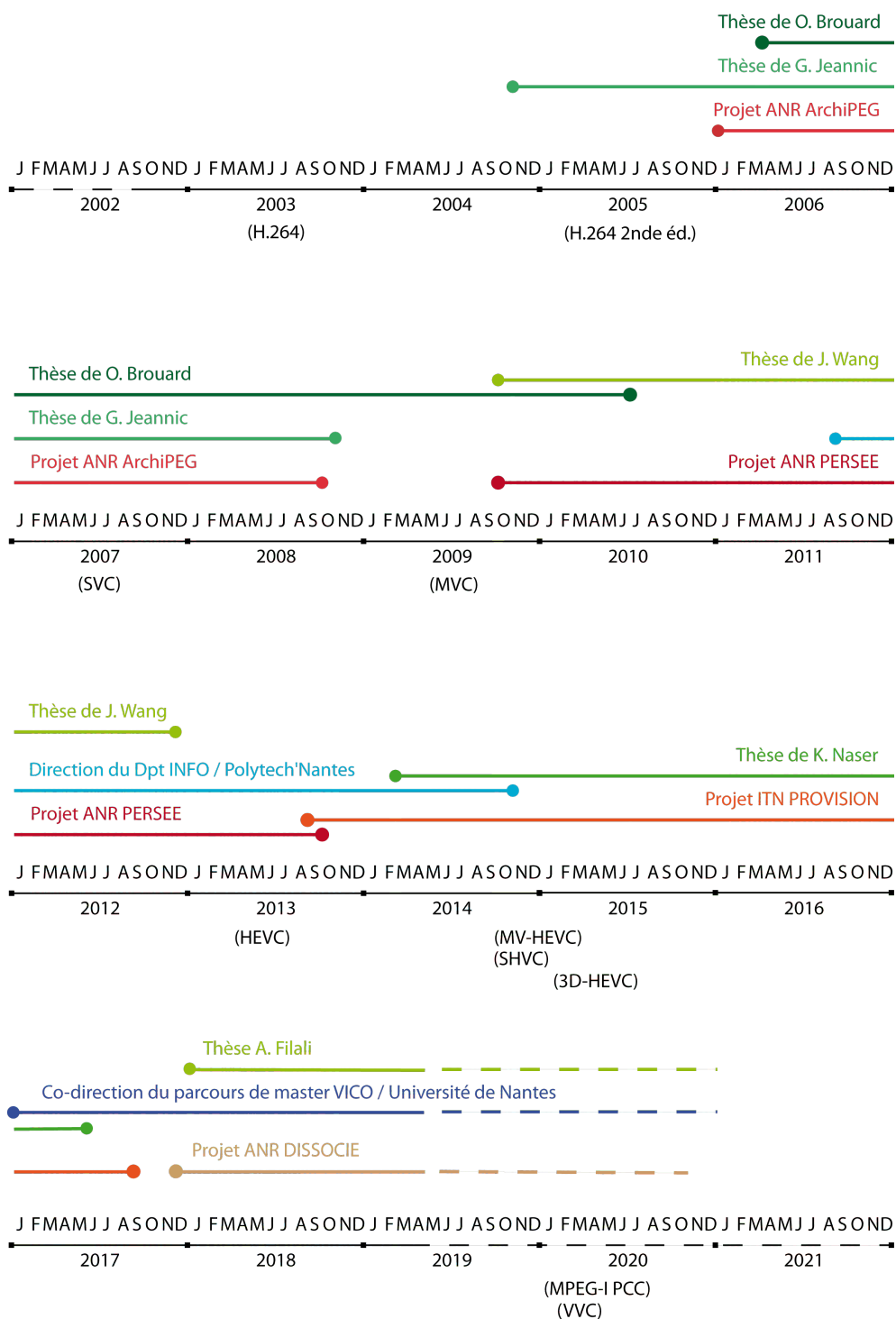


FIGURE 2.2 – Vue chronologique des principaux évènements décrits dans le mémoire et en particulier ceux relatifs à mes contributions au codage perceptuel.

Première partie

**Contributions aux projets de  
recherche**

---

## Introduction

La création de l'ANR<sup>11</sup> « Agence Nationale de la Recherche » en 2005, avec pour mission de mettre en place le financement de la recherche sur projets en France, a profondément impacté l'environnement et le mode de travail des équipes. Sachant qu'ensuite toute une hiérarchie de programmes (du niveau régional à celui européen) s'est progressivement mise en place pour assurer la sélection, le financement et le suivi de projets de recherche<sup>12</sup>. En quelques années les laboratoires (notamment ceux comme l'IRCCyN du département STIC « Sciences et Technologies de l'Information et de la Communication » du CNRS) qui se sont adaptés, ont changé de ressources budgétaires pour fonctionner et investir. Les dotations récurrentes stagnent (voire régressent), les contrats sont devenus la manne principale. Par exemple l'IRCCyN a vu la part de ses dotations récurrentes passer de 25% de son budget en 2003 à 12% en 2007. Au niveau de l'équipe IVC, dès 2008 les contrats représentaient 90% des financements (hors salaires des personnels permanents), ce taux s'est maintenu et renforcé jusqu'en 2017. Pour l'équipe IPI du LS2N la situation a évolué favorablement, mais la part de financement des projets demeure toujours incontournable.

Étant donné qu'un projet dure en moyenne 3 ans, il faut les enchaîner pour maintenir la structure de travail car les contrats financent des salaires (par exemple : doctorants, post-doctorants, stagiaires, ingénieurs de recherche, gestionnaire de projet), des participations aux activités de la recherche (réunions, conférences, etc.), du matériel et le maintien de plateformes expérimentales. Actuellement, l'équipe IPI assure en moyenne une participation à une dizaine de projets en parallèle, bien entendu le panel est large avec des projets aux périmètres très différents (européen, FUI, ANR, OSEO/PBI, Cifre, Pays de Loire, etc.), variant notamment au niveau de leur objectif (recherche fondamentale en amont *vs.* démonstrateur industrialisable) et de leur partenariat (national *vs.* international, académique *vs.* entreprise, grand groupe *vs.* PME, etc.).

L'enseignant-chercheur a aussi vu sa mission impactée. De façon cyclique il est amené à répondre (avec un réseau sélectionné de partenaires) à des appels à projets, à ce niveau en plus d'être un scientifique, il doit aussi être un gestionnaire capable d'estimer des coûts financiers et humains. Il faut souvent répondre à plusieurs appels à projets afin d'en obtenir un (par exemple pour l'ANR, le taux de succès tous programmes confondus varie d'une année sur l'autre : de 25% en 2005, il tombe à 20% en 2012, à 8,5% en 2014, ... il est remonté à 15,1% en 2018). Lorsque le projet est sélectionné, l'enseignant-chercheur participe aussitôt à la mise en place des étapes préalables à son exécution : embauches (doctorants, post-doctorants, ingénieurs), démarches administratives multiples (déclarations aux services concernés de l'université et du laboratoire, négociation du salaire, obtention de visas, etc.), rédaction et/ou relecture de documents juridiques (accords de confidentialité, propriété intellectuelle). Pendant la durée du projet, l'enseignant-

---

11. Site web de l'ANR : [www.agence-nationale-recherche.fr](http://www.agence-nationale-recherche.fr)

12. Il y a eu en parallèle un véritable bouleversement du paysage universitaire (cf. loi de programmation de la recherche en 2006, et loi dite « LRU » en 2007) qui a incité les établissements publics à se développer. Cet aspect n'est pas développé ici.

---

chercheur contribue à son pilotage jalonné de réunions de suivi, de rédactions de rapports (avancée des travaux, justification de l'utilisation des fonds), de livraisons à faire suivant un échéancier, de maintien d'outils (site web, listes de diffusion, serveurs pour l'hébergement et l'échange des données, etc.), d'actions de communication vers le grand public ou la communauté scientifique, d'audits intermédiaires et de celui final. Il doit manager une équipe de doctorants et d'ingénieurs. L'encadrement de la thèse est aussi influencé, car il faut conduire le doctorant à appréhender le contexte et les enjeux du projet, et lui faire prendre part responsabilisé aux tâches qui incombent (rédaction de rapports, participation aux audits, actions de communication, etc.).

La recherche sur projets représente donc plus de travail pour les équipes de recherche, l'aspect positif est son effet catalyseur élargissant le spectre des partenariats (notamment à l'international) et le nombre des publications, et une augmentation de moyens financiers (embauches, réalisation et maintien de plateformes expérimentales). Dans le domaine du codage vidéo où la recherche vise à faire évoluer les standards, ces projets offrent l'opportunité de construire des consortiums solides s'organisant pour trouver et défendre de nouveaux procédés. Cette partie du mémoire décrit mon parcours au sein de projets de recherche qui se sont enchaînés depuis 2005 et ont tous contribué au codage vidéo perceptuel. Nous verrons comment ils se caractérisent (contexte, objectifs, réalisations), et je présenterai mes contributions (responsabilités organisationnelles et scientifiques) à ces projets, ainsi que mon bilan personnel.



## Chapitre 3

# Le projet ANR ArchiPEG

### 3.1 Contexte et objectifs

Le réseau RIAM pour la « Recherche et l’Innovation en Audiovisuel et Multimédia » soutient les projets de Recherche & Développement des PME des secteurs de l’audiovisuel et du multimédia. Il réunit depuis 2001 le CNC « Centre national du cinéma et de l’image animée » et les ministères délégués à la recherche et l’industrie. Il a été associé la 1<sup>re</sup> fois à l’ANR en 2005 pour un appel à projets.

VITEC<sup>1</sup>, se définit alors comme une PME leader dans le « domaine vidéo MPEG » offrant des solutions matérielles et logicielles pour l’encodage/décodage, l’édition ou le multiplexage de vidéos. ENVIVIO est aussi une PME leader dans le domaine des solutions complètes de « broadcast » et de « streaming » basées MPEG-4. Les 3 partenaires (VITEC, ENVIVIO et l’IRCCyN) se connaissaient, car IVC a expertisé le codeur MPEG-2 de VITEC (contrat en février 2005), et VITEC et ENVIVIO ont collaboré au sein du projet RIAM COPARO dont le principal objectif était la mise au point d’un codeur temps réel H.264/AVC au format SD « Standard Definition ». S’intéressant aux incidences de l’introduction du format vidéo HD dans les nouveaux services de diffusion [395, 403, 424], le futur consortium pouvait faire le constat de l’émergence de nouveaux médias et services (la TNT « Télévision Numérique Terrestre », les réseaux mobiles 3G, la VoD « Video-on-Demand », la TV HD, la TV sur terminaux mobiles, le cinéma numérique 2K et 4K), mais aussi de la grande complexité du standard H.264/AVC proposé pour encoder le format HD [404, 468] (la TV HD représente 5 fois plus de pixels à traiter que la TV SD avec une complexité de traitement entre 6 à 10 fois supérieure [421], voir aussi la figure 3.1).

Fin 2005, VITEC (porteur) et ses 2 partenaires répondent à l’appel du RIAM dans le cadre de la thématique s’intéressant aux « nouveaux modes de distribution de contenus audiovisuels et multimédia », et définissent trois objectifs à leur projet nommé ArchiPEG

---

1. Site web de la société VITEC : [www.vitec.com](http://www.vitec.com)

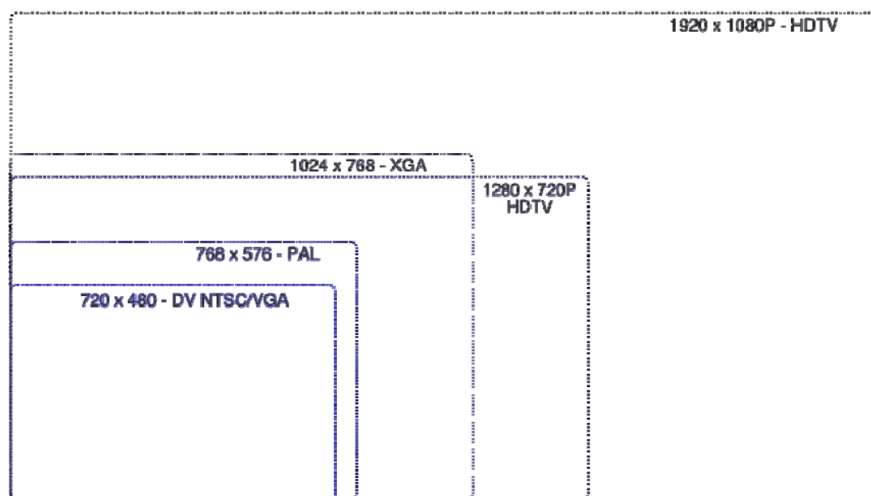


FIGURE 3.1 – Différentes résolutions existantes en SD et HD TV (extrait de [315]).

pour « Architecture modulaire, Programmable, Evolutive, Graduable pour les applications en vidéo numérique de très hautes performances » :

1. La conception et la mise au point d'une nouvelle architecture (matérielle et logicielle) de traitement de la vidéo capable d'intégrer H.264/AVC et de répondre aux besoins de la TV HD (formats 720p, 1080i, 1080p) et du cinéma numérique (2K et 4K). Cette architecture devra être modulaire (remplacement aisé de fonctions), programmable (flexibilité matérielle et logicielle), évolutive (intégration des futurs matériels et technologies), graduable (indépendance des définitions des images), et ouverte (intégration de cartes matérielles standards).
2. La compression H.264/AVC en temps réel du format vidéo HD (cette application n'avait pas encore en 2005 de solution industrielle).
3. La mise au point d'un algorithme de pré-traitement afin de caractériser le mouvement au sein de la vidéo, et piloter les choix des meilleurs modes de prédiction à l'encodage.

## 3.2 Organisation

ArchiPEG est un projet pré-compétitif de 24 mois, il vise donc le développement technique et la croissance des PME participantes. Exactement il a duré du 1<sup>er</sup> janvier 2006 au 12 octobre 2008 car une prolongation de 10 mois sera accordée. L'aide globale allouée par l'ANR est de 1810 k€ pour 172 homme-mois, incluant l'enveloppe de l'IRC-CyN de 200 k€ pour 48 homme-mois : 24 homme-mois d'un ingénieur d'études et 24

homme-mois d'un doctorant (il faut noter qu'en sus un contrat sera passé avec VITEC pour le financement d'une 3<sup>e</sup> année de thèse). Le travail est organisé en 5 sous-projets :

**Sous-projet 1** Coordination.

**Sous-projet 2** Conception et mise au point d'une plateforme hardware modulaire de traitement temps réel.

**Sous-projet 3** Développement de logiciels d'encodage vidéo haute définition.

**Sous-projet 4** Pré-analyse et conditionnement du flux vidéo.

**Sous-projet 5** Validation des résultats, c'est-à-dire produire :

- Une chaîne de traitement permettant de vérifier l'encodage temps réel au standard H264/AVC en « HighProfile » d'un flux vidéo HD (au moins 2 formats 720p et 1080i).
- Une qualité visuelle du résultat décodé au moins égale à celle obtenue en encodant/décodant au standard MPEG-2 HD et pour un débit réduit de 30 à 50%.
- Plusieurs implémentations d'un même algorithme sur la plateforme (avoir différentes ressources de calcul).

Les partenaires se complètent car VITEC doit coordonner le projet, et concevoir et réaliser l'architecture matérielle, ENVIVIO doit développer l'algorithme de codage temps réel, et enfin l'IRCCyN (l'équipe IVC) se charge des algorithmes pour le filtrage et la pré-analyse des flux vidéo afin de déterminer, en amont du codeur, une stratégie pour l'encodage du flux vidéo HD au format H.264/AVC. Enfin tous les partenaires doivent contribuer à la validation des résultats.

### 3.3 Travaux menés et résultats

VITEC conçoit et réalise début 2006 la plateforme matérielle de traitement montrée à la figure 3.2. Celle-ci accueille plusieurs noeuds de calcul (chaque noeud étant une carte au format PMC comportant 5 DSP et 1 FPGA, voir la figure 3.3) tout en assurant des communications à suffisamment haut débit pour que leur impact soit transparent sur le temps de calcul.

Après avoir réalisé l'architecture matérielle, VITEC a aussi assuré le développement de l'algorithme de codage car ENVIVIO ne sera présent qu'au dépôt du projet. Un premier prototype a été réalisé fin 2006, incluant une librairie de base permettant d'assurer des services de communication et de synchronisation de tâches entre processeurs de la plateforme, et un encodage simplifié de flux vidéo HD. Le travail porte ensuite sur l'extension à la HD du logiciel d'encodage H.264/AVC simple définition fonctionnant en



FIGURE 3.2 – La plateforme matérielle ArchiPEG (extrait de [304]).

temps réel sur la gamme des cartes VP3 de VITEC. En 2007, un encodage H.264/AVC sur flux HD est réalisé. Plusieurs variantes sont proposées, dont une optimisée pour fonctionner en environnement DSP et permettant d'atteindre les performances temps réel sur la plateforme matérielle ArchiPEG. L'encodage H.264/AVC est alors réalisé au profil MP « Main Profile », et il y a parallélisation des traitements en découpant les images jusqu'à 8 slices. Cette application a permis de démontrer les performances de la plateforme aussi bien en termes de puissance de calcul que de capacité de communication interne ou externe. Par exemple, la comparaison des encodages temps réel en flux H.264/AVC *vs.* MPEG-2 de vidéos HD, donne un gain moyen en débit de 45% (à qualité objective égale).

L'IRCCyN a réalisé un outil pour la pré-analyse de la vidéo HD avant son encodage en flux H.264/AVC. Ces travaux [38, 19, 40, 20, 21] seront détaillés dans le chapitre 7 de la partie II, ils visent à encoder les plans vidéo en tenant compte de leur caractéristiques perceptuelles. Nous rappelons brièvement ici que l'outil est placé en amont du logiciel de codage, il découpe la vidéo HD en plans homogènes et combine :

- Un traitement intra-segment temporel qui segmente en objets cohérents vis-à-vis du codage. La segmentation est d'abord basée mouvement (tube spatio-temporel) puis basée multi-critères (couleurs et textures, via une approche Markovienne).
- Un traitement inter-segments temporels pour suivre les objets sur plusieurs segments contigus. Cela conduit au calcul d'une carte de saillance, qualifiant l'importance psycho-visuelle des objets, afin de piloter leur encodage (choix des modes de prédiction et quantification).

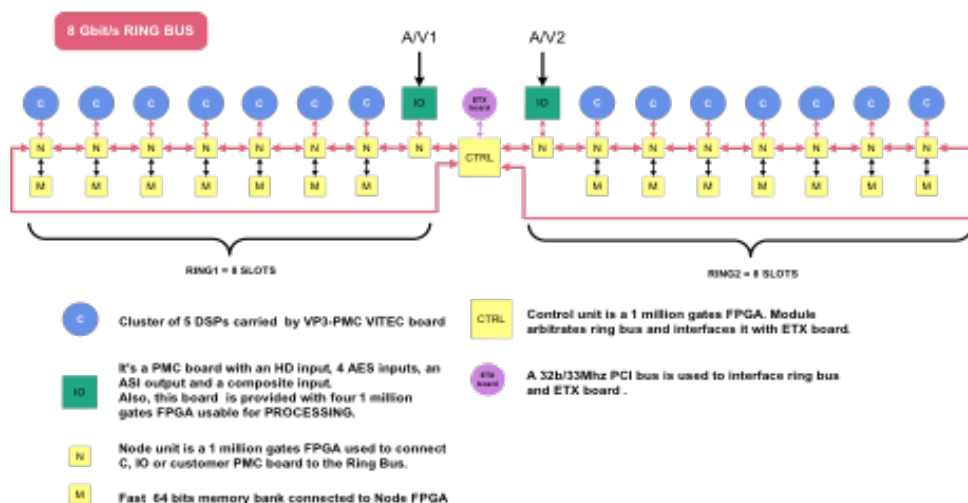


FIGURE 3.3 – Architecture générale de la plate-forme matérielle ArchiPEG avec son système d'interconnexion à double anneaux (extrait de [304]).

VITEC a intensifié sa coopération avec l'IRCCyN lors des travaux sur l'encodage vidéo basé sur critères perceptuels. Pour la mesure de qualité des vidéos décodées, le budget ne permettant pas une campagne de tests subjectifs, VITEC a retenu l'outil VQA « Video Quality Monitor » d'AccepTV<sup>2</sup>.

L'arrivée sur le marché des processeurs multi-cœurs massifs entrant en compétition directe avec les plateformes matérielles de types ArchiPEG. Il est apparu nécessaire pour VITEC de réorienter l'objectif du dernier sous-projet pour mieux préparer l'avenir. Dans cette perspective, il a été pris contact avec l'IETR<sup>3</sup> (équipe IMG<sup>4</sup> « Image » à Rennes) ayant une connaissance certaine sur ce nouveau type d'architecture grâce à leur expérience sous SynDEx [526] « Synchronized Distributed Executive ». Cet environnement, conforme à la méthodologie AAA « Adequation Algorithm Architecture » permet un prototypage rapide et une implémentation optimisée d'applications temps réel distribuées. Le logiciel d'encodage a donc été replacé dans la plateforme de simulation SynDEx afin de tester différents modes de parallélisation, sur des architectures à mémoire distribuée DRAM « Distributed Random Access Machines » (communication entre processeurs par passage de messages) ou à mémoire partagée PRAM « Parallel Random Access Machines » (communication par mémoire partagée), et les comparer à des implémentations temps réel. Les résultats des travaux ont montré que le prototypage automatique offrait des performances proches d'une parallélisation manuelle et qu'il était possible de trouver une configuration optimale fonction de l'architecture de traitement visée. Ces résultats très encourageants ont montré que les efforts de développement logiciel réalisés dans le cadre du projet pouvaient être directement réutilisés sur de nouvelles architectures de

2. AccepTV est un essaiage de l'équipe IVC, son site web : [www.acceptv.com](http://www.acceptv.com)

3. Site web de l'IETR : [www.ietr.fr](http://www.ietr.fr)

4. Site web de l'équipe IMG : [www.ietr.fr/spip.php?article1594](http://www.ietr.fr/spip.php?article1594)

traitement parallèle. Les principaux résultats issus du projet sont donc :

- Une plateforme multiprocesseur permettant de réaliser des traitements temps réel sur flux vidéo HD.
- Un logiciel d’encodage vidéo avancé écrit en langage C et optimisé pour 2 types de processeurs cibles (processeur généraliste multi-cœurs Intel, processeur de traitement de signal Texas Instruments).
- Un logiciel de pré-analyse et de conditionnement de flux vidéo permettant de paramétrer l’encodage selon des critères perceptuels.
- Une démarche efficace pour porter rapidement sur architecture de calcul parallèle le logiciel d’encodage vidéo avancé.
- Des implémentations fonctionnant en temps réel sur architecture de traitement parallèle.

### 3.4 Contributions et bilan personnels

Mes contributions d’ordre scientifique aux sous-projets 4 et 5 sont décrites dans le chapitre 7 (partie II). Il faut remarquer que sur le plan scientifique l’approche est particulièrement ambitieuse car elle est globale, les paramètres issus de l’analyse du contenu du plan vidéo servant directement à piloter perceptuellement le codage en exploitant la panoplie d’outils qu’offre la norme de codage H.264/AVC.

Mes contributions ont aussi été d’ordre organisationnel, on retrouve la liste de tâches qui impactent la mission de l’enseignant-chercheur et que j’ai répertoriées dans l’introduction I de la partie I. Plus spécifiquement les démarches administratives au début du projet ArchiPEG ont été lourdes car au 1<sup>er</sup> semestre 2006, la mise en place des projets ANR étant nouvelle, l’université n’offre pas encore tous les services appropriés. Par exemple il n’y a pas encore possibilité de recruter directement en CDD (c’est la filiale CAPACITÉS<sup>5</sup> qui embauche, cela implique des démarches supplémentaires pour établir les contrats... et cela a un coût), ou bien encore il faut négocier directement avec les partenaires l’accord de consortium. Il faut ajouter que pendant la période nous avons aussi établi avec VITEC le contrat pour le financement de la 3<sup>e</sup> année de thèse. Ces démarches et l’attente de leur aboutissement ont retardé de 6 mois le début de la thèse (et justifié la renégociation fin 2007 d’un avenant pour la prolongation du projet). Après son lancement, ArchiPEG a eu un déroulement quasi-linéaire avec, des rapports semestriels d’activités, des livraisons (nous avons produit quatre livrables [274, 275, 312, 339]), deux audits (un à mi-parcours et le final), des réunions de travail régulières (quatre par an et en alternance chez les partenaires), les états et justifications des dépenses. Nous avons

---

5. CAPACITÉS est la filiale privée de l’université de Nantes dédiée à la valorisation de la recherche, son site web est : [www.capacites.fr](http://www.capacites.fr)

aussi communiqué de façon spécifique sur le projet en participant au « Grand Colloque STIC » organisé par l'ANR à Lyon en 2008.

ArchiPEG a été ma 1<sup>re</sup> expérience d'un projet d'envergure auquel j'ai participé avec des responsabilités variées et de bout en bout. Cela a aussi été mon 1<sup>er</sup> long partenariat avec une PME dans un contexte de projet pré-compétitif. Si les contraintes industrielles et économiques ont eu des péripéties sur la vie du projet (cf. ré-orientation du sous-projet 5), elles ont eu peu d'incidence sur les orientations de notre recherche. ArchiPEG a aussi procuré plus de moyens pour accompagner le travail de thèse, avec l'embauche de l'ingénieur d'études qui est intervenu pour développer les procédés et les implémenter dans le logiciel de référence. ArchiPEG a donc montré l'intérêt d'une recherche sur projets en codage vidéo, les clés du succès étant certainement dans la qualité du partenariat et sa complémentarité.

## Chapitre 4

# Le projet ANR PERSEE

### 4.1 Contexte et objectifs

Début 2009, l'émergence des formats vidéo HD s'est accompagnée de nouveaux services immersifs (affichage panoramique HDTV et celui interactif avec la vidéo omnidirectionnelle ou FVV « Free Viewpoint Video ») ou la 3DTV (avec la vidéo stéréo et celle multi-vues). Les efforts de recherche en codage vidéo visent alors à améliorer la compression de séquences multi-vues, et une activité de standardisation 3DVC « 3D Video Coding » est organisée depuis 2007 par le JVT [238, 363]. Un standard MVC « MultiView Coding » [285] doit paraître comme une extension de H.264/AVC. Cependant le rendu de la qualité visuelle est un problème plus crucial en vidéo 3D qu'en vidéo classique 2D, et les 1<sup>ers</sup> résultats annoncent que ce standard initial aura des gains de codage insuffisants et souffre de limitations fonctionnelles lorsque qu'il s'agit de rendre avec qualité (coté récepteur) des points de vue virtuels. Les espoirs portent sur la future norme de codage vidéo HEVC [164] encore en gestation.

Les quatre laboratoires partenaires (IRCCyN, INRIA Rennes<sup>1</sup>, IETR Rennes et LTCI Télécom ParisTech<sup>2</sup>) font le constat qu'il faut mettre en avant et améliorer la qualité visuelle des vidéos décodées 2D (contenu classique) et 3D (synthèse coté récepteur de points de vue virtuels). L'objectif d'une représentation perceptuelle adaptée au 2D et à la 3D, et son exploitation pour piloter le codage vidéo, s'appuie sur différents modèles prometteurs et sur des techniques émergentes mais qui nécessitent encore des recherches plus approfondies. Ces travaux qui retiennent l'attention du consortium sont les modèles perceptuels et de saillance visuelle [267, 321, 351, 367, 389, 390, 420, 423, 424, 436, 538, 543, 569, 586, 611], les métriques de qualité [270, 292, 309, 362, 426, 435], la représentation et l'analyse des images par ondelettes [343, 383], l'analyse/synthèse de textures spatio-temporelles [300, 386, 442], l'optimisation débit-distorsion [288, 473, 546], le codage robuste par description multiple [330, 501, 503, 515, 531, 541, 555, 589, 631], et l'estimation de la disparité des images stéréo et multi-vues [488].

---

1. Le site web de l'INRIA Rennes : [www.inria.fr/centre/rennes](http://www.inria.fr/centre/rennes)

2. Le site web du LTCI : [lcti.telecom-paristech.fr](http://lcti.telecom-paristech.fr)



Les 4 partenaires s'associent pour répondre au « Programme ANR BLANC édition 2009 » (le CSD « Comité scientifique disciplinaire » ciblé étant celui « Sciences et technologies de l'information et de la communication »), et définissent le projet PERSEE<sup>3</sup> « PERceptual Scheme for 2D and 3D vidE(E)o coding » soit « Schémas perceptuels & codage vidéo 2D et 3D ». Les programmes « blanc » de l'ANR sont des appels à projets non thématiques, ils ont pour but de donner une impulsion à des projets scientifiques ambitieux qui se positionnent dans la compétition internationale, et qui présentent des objectifs originaux en rupture avec les itinéraires de recherche traditionnels. Pour les partenaires académiques s'est l'opportunité de développer ensemble une recherche en amont, car les retombées attendues du projet sont des thèses, des publications et des brevets.

## 4.2 Organisation

Le projet PERSEE a duré du 1<sup>er</sup> octobre 2009 au 30 septembre 2013 (sa durée initiale de 36 mois ayant été prolongée de 12 mois). L'aide globale allouée par l'ANR est de 724 k€ dont 186 k€ pour l'IRCCyN, ce programme limite le montant ETP « équivalents temps plein » financés à 72 homme-mois et typiquement pour l'embauche de doctorants. Pour PERSEE un gros travail d'ingénierie est à prévoir pour développer et intégrer les nouveaux outils<sup>4</sup>, la solution (avec l'accord de l'ANR) a été l'augmentation du nombre homme-mois à budget constant afin de recruter des ingénieurs, et le financement partiel ou total de thèses sur fonds propres des équipes. Pour atteindre ses objectifs, PERSEE s'est organisé autour de 6 grandes tâches (plus une de coordination, voir aussi la figure 4.1) :

**Tâche 0** Coordination des aspects techniques, administratifs et financiers.

**Tâche 1** Identification et amélioration des modèles perceptuels des contenus 2D & 3D, puis leur utilisation par les autres tâches.

**Tâche 2** Développement d'approches pour l'analyse des caractéristiques (géométriques, structurelles et texturelles) des contenus visuels, et la synthèse des textures spatio-temporelles.

**Tâche 3** Développement de méthodes pour la représentation et le codage des signaux vidéo 2D, en tenant compte de la qualité perceptuelle des signaux reconstruits (lien avec la tâche 1) et de la compatibilité souhaitée avec la représentation des vidéos 3D.

---

3. Le site web du projet PERSEE : [persee.irccyn.ec-nantes.fr](http://persee.irccyn.ec-nantes.fr)

4. En 2009, les 4 partenaires soumettent exactement 2 projets complémentaires à l'ANR. L'un 3DnV « Vidéos naturelles en 3D » répond à l'appel thématique CONTINT « Contenus et Interactions » et vise à développer avec des PME (dont VITEC) une solution dans le cadre HEVC pour améliorer l'expérience 3D ressentie par le spectateur, un gros travail d'ingénierie est prévu. L'autre projet soumis est PERSEE, mais seul ce dernier sera retenu par l'ANR.

**Tâche 4** Développement de techniques pour la représentation et le codage des signaux vidéo multi-vues (3D) en tenant compte des résultats des tâches 1 (adaptation au codage perceptuel) et 3 (compatibilité avec la représentation 2D).

**Tâche 5** Développement d'un nouveau codec intégrant les contributions des partenaires issues des autres tâches. La base initiale de cette plateforme logicielle commune est un codec 2D de l'état de l'art, qui doit ensuite évoluer en fonction des contributions. Les vidéos codées/décodées à l'aide des configurations les plus prometteuses de la plateforme seront retenues pour être évaluées à la tâche 6.

**Tâche 6** Evaluation subjective des performances visuelles et perceptuelles des outils développés et du codec.

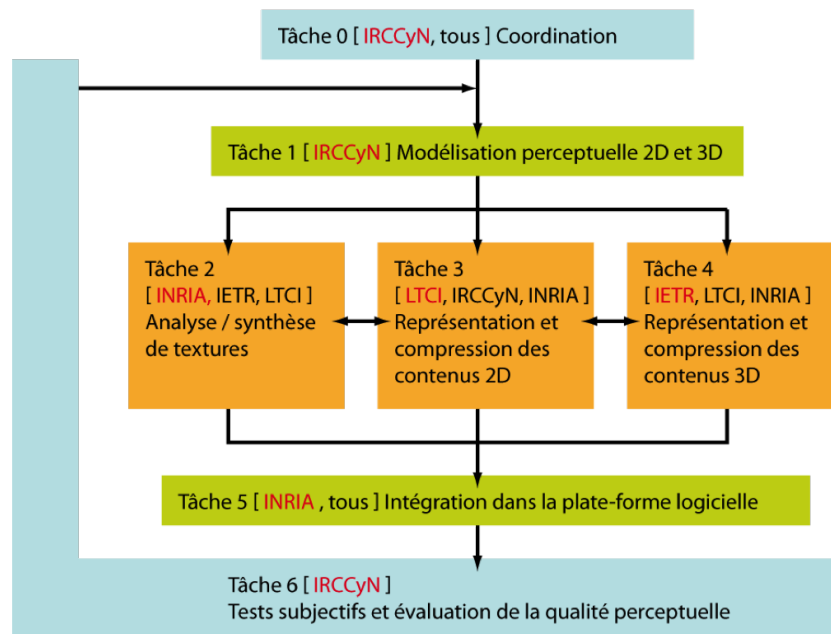


FIGURE 4.1 – L'organisation en tâches de PERSEE avec leurs contributeurs, le partenaire responsable de la tâche est indiqué en rouge.

Les équipes de recherche partenaires se connaissent et se complètent, chacune doit contribuer dans son domaine d'expertise. La coordination est assurée par l'IRCCyN, et précisément j'ai eu cette responsabilité (mon bilan est présenté au paragraphe 4.4). L'équipe IVC/IRCCyN doit aussi principalement concevoir les modèles perceptuels pour la représentation et le codage des signaux vidéo 2D&3D. IVC doit également conduire, via ses plateformes expérimentales, les tests subjectifs nécessaires pour évaluer la qualité visuelle des vidéos 2D&3D codées/décodées par le consortium. L'équipe TEMICS/INRIA<sup>5</sup> doit travailler à l'analyse-synthèse des textures via de nouvelles approches (ondelettes,

5. Le site web de l'équipe TEMICS : [www.inria.fr/equipes/temics](http://www.inria.fr/equipes/temics)

représentations creuses, etc.) et pour des contenus vidéo aussi bien mono-vue que multi-vues (et multi-vues plus profondeur). TEMICS doit aussi contribuer avec IVC au codage perceptuel. Enfin TEMICS doit mener la construction de la plateforme logicielle commune pour évaluer et valider les outils et approches développés par le projet. L'équipe IMG/IETR doit principalement contribuer à la représentation et la compression des contenus 3D, tout en veillant à leur compatibilité avec celles des contenus 2D. IMG doit collaborer plus spécifiquement avec IVC, pour tenir compte des résultats d'évaluations perceptuelles dans ses algorithmes, et avec le LTCl qui doit fournir des données (disparités estimées) améliorant le traitement des données multi-vues. Finalement le Groupe Multimédia<sup>6</sup> du LTCl, doit prendre en charge la représentation et la codage des contenus 2D avec des approches innovantes (meilleure exploitation de l'information de mouvement, nouvelles techniques de représentation et de transformation de l'information 2D, etc.).

### 4.3 Travaux menés et résultats

Pour la tâche 1, PERSEE s'est intéressé à la modélisation de l'attention visuelle qui est un mécanisme mis en œuvre par le SVH afin de réduire la quantité d'information que le cerveau a besoin de traiter pour appréhender le contenu d'une scène (je détaille dans le chapitre 8 de la partie II, les travaux menés par l'IRCCyN dans le cadre de la thèse de June Wang). Plusieurs études ont été menées, la première concerne la modélisation de l'attention visuelle en image fixe avec l'analyse de la fiabilité de cartes de densité de fixation issues de différentes bases de données oculométriques [242]. Les autres études sont relatives à la modélisation de l'attention visuelle dans des conditions de visualisation 3D [5]. Pour faire face au manque de vérité de terrain exploitable il a d'abord été créé, via des expérimentations oculométriques binoculaires, une nouvelle base de données avec des images stéréoscopiques 3D. L'impact du « biais de profondeur » sur l'attention visuelle dans des conditions de visualisations 3D sur écran plat stéréoscopique, a été quantifié [3, 18]. Des modèles de l'attention visuelle 3D ont été développés [2, 144, 146, 192], ils reposent sur le contraste de profondeur. Le « biais central » en condition de visualisation 2D ou 3D, a aussi été analysé, puis intégré aux modèles de l'attention visuelle 3D. La relation entre flou et disparité binoculaire, a été analysée [197, 34, 17] et il a été montré comment une technique de floutage peut améliorer la qualité d'expérience de la TV3D. Enfin une étude a pris en compte la dimension temporelle [194, 195] pour l'intégrer à la modélisation de l'attention visuelle 2D ou 3D.

Pour la tâche 2, les outils pour l'analyse et la synthèse de texture, ont été définis pour trois contextes d'applications. La 1<sup>re</sup> application concerne la compression d'images et plus particulièrement les méthodes de prédiction. Les études ont été faites dans un contexte de compression H.264. Quatre méthodes ont été développées et testées [129] : TM (Template matching), ATM (Average Template Matching), LLE (Locally Linear Embedding), NMF (Non-negative Matrix Factorization). La 2<sup>e</sup> application porte sur des approches de rebouchage [95, 118, 154] (ou « inpainting ») reposant sur des techniques de

---

6. Le site web du Groupe Multimédia : [www.tsi.telecom-paristech.fr/mm](http://www.tsi.telecom-paristech.fr/mm)

copier/coller, encore appelées synthèse de texture basée sur l'exemple. L'approche scientifique considérée a été d'améliorer un algorithme existant. Cette méthode est constituée de deux étapes principales : un calcul de priorité sur la ligne de front (ligne séparant la zone connue de la zone à remplir) et une recherche de candidat. Ces deux étapes ont été améliorées en considérant des tenseurs de structures et des approches de type  $K$ -NN respectivement pour la définition de la priorité et de la recherche de candidat (voir aussi la figure 4.2). La 3<sup>e</sup> application concerne la synthèse de vue virtuelle. Il s'agit ici de créer un point de vue virtuelle en utilisant soit une vue existante soit un ensemble de vues. La problématique principale rencontrée concerne la projection d'une vue existante dans un nouveau référentiel et la gestion des zones découvertes. Un algorithme effectuant la projection et le traitant des zones découvertes (ainsi que les craquelures) a été proposé [132, 168, 169, 185, 193, 204].

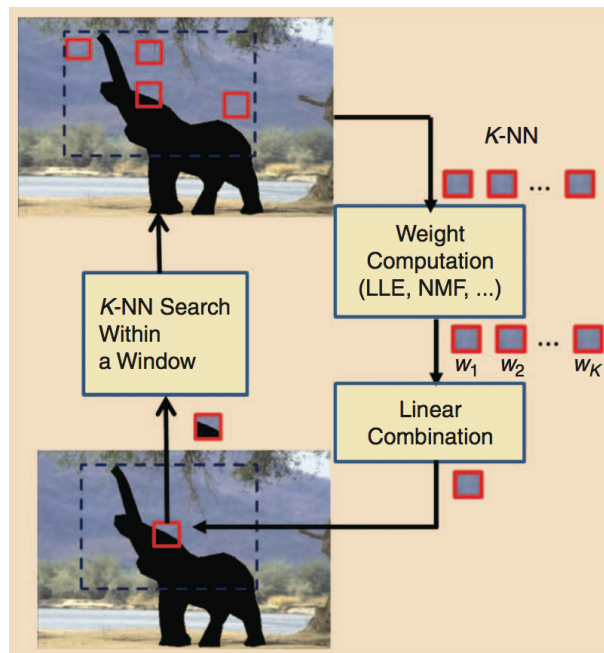


FIGURE 4.2 – Inpainting : principe des approches de type  $K$ -NN (d'après [95]).

Pour la représentation et le codage des contenus 2D de la tâche 3, la méthodologie employée a consisté à concevoir des modèles pertinents du signal à coder. Un 1<sup>er</sup> aspect pris en compte a été la représentation du mouvement, qui est une des sources fondamentales d'information dans la vidéo 2D. Des algorithmes d'estimation du mouvement « denses » (c'est-à-dire déterminant la trajectoire de chaque pixel) ont été conçus et implémentés [219]. Ils permettent de s'affranchir des artéfacts dus à la représentation traditionnelle par blocs. Des réductions de débit (à qualité égale) de l'ordre de 1% ont été obtenues, ce qui est considéré comme un bon gain par rapport aux normes. Un 2<sup>nd</sup> axe a regardé de nouvelles techniques de transformée linéaire [220, 227]. Cette dernière est un outil fondamental en compression d'images et de vidéos, car elle permet de concentrer

l'information pertinente sur peu d'échantillons. Toutefois, les performances des techniques de transformée classiques se détériorent à proximité des contours des objets ou des structures avec orientations spatiales spécifiques. De nouvelles transformées ont donc été conçues et implémentées afin de prendre en compte ces caractéristiques fondamentales des images ; les résultats ont montré une amélioration objective et perceptuelle de la qualité. D'autres contributions sur la compression 2D ont regardé la représentation progressive de la vidéo et le codage par exemplaires.

Pour la tâche 4, des méthodes de représentation et de codage de vidéos 3D ont été proposées au-delà des standards existants (c'est-à-dire MPEG 3DMC pour la stéréo et H.264/MVC pour les séquences multi-vues). La compression de données MVD (multi-vues plus profondeur) a été étudiée en particulier : à partir des représentations LDI et LDV (Layered Depth Image / Video), des extensions I-LDI et O-LDI ont été proposées et leurs performances évaluées pour la compression dans le cadre du FVV [185, 188, 189]. Plusieurs algorithmes pour la compression de cartes de profondeur ont été développés [99, 108, 111, 112, 124, 125, 138, 140, 147, 170, 176, 181, 183, 215, 216], avec des approches destinées à favoriser la qualité perceptuelle lors de la synthèse de vues : approche contours et courbes élastiques, approche quad-tree (LAR codec), approche basée sur le paradigme « Don't Care Regions », et approche par « Depth Fading ». Ces méthodes ont été conjointement évaluées en coopération avec la tâche 6, dans des tests psychovisuels impliquant tous les partenaires du projet. Plusieurs algorithmes ont été intégrés dans les schémas en cours de standardisation 3D-HEVC [123, 158], et les plus performants proposés en normalisation : la modification de la liste MERGE des candidats, une modification de la dérivation du vecteur de disparité, et l'utilisation de la JPF (Joint Projection Filling) pour la prédiction inter-vues.

Pour la tâche 5 d'intégration dans la plateforme logicielle, les études menées se sont concrétisées par des réalisations logicielles de deux natures : des applications autonomes (ou stand-alone), et des intégrations au sein de logiciels de référence mis à la disposition de la communauté par les organismes de normalisation. Les applications « stand-alone » concernent le codage des cartes de profondeurs (« Don't Care Region », codage par diffusion préservant les contours et codage basé LAR), les modèles d'attention visuelle et des techniques d'inpainting. Plusieurs approches étudiées dans les tâches précédentes ont fait l'objet d'une intégration dans les logiciels de références de codage vidéo 2D et 3D. Pour la 2D, le « Weighed Template Matching » (WTM) a été intégré au « test model HM » de HEVC puis évalué. Les travaux liés à la 3D ont quant à eux été intégrés dans le « test model JM » de H.264 pour le DCR et le HTM, le « test model » basé HEVC et dédié à la 3D pour les autres travaux. En particulier, cela a concerné les « Layer Depth Image » (LDI), et le « Joint Projection Filling » (JPF). Deux autres intégrations dans le HTM ont fait l'objet d'une adoption en normalisation. Il s'agit de l'ajout de candidats dérivés de vecteurs de disparités dans la liste des prédicteurs de mouvement, et du codage de carte de profondeur hérité de la décomposition en quadtree des blocs de la texture. Elles font désormais partie du standard développé pour la vidéo 3D par le groupe JCT-3V. Parallèlement à ces logiciels, une base de données de cartes de profondeurs compressées et associées à des notes de qualités a été développée. Elle

permet d'aider à l'évaluation de métriques de qualité.

Pour la tâche 6, des campagnes de tests subjectifs (tests de qualité, expérimentations autour de la perception visuelle psychophysique) ont été régulièrement menées en lien avec les travaux de recherche menés par les partenaires. Cela a mobilisé un ensemble de moyens : salles d'expérimentations normalisées, matériel spécialisé (serveur haut débit, systèmes professionnels d'affichage HD&3D, oculomètres, bases de données, etc.), panels d'observateurs, et maîtrise de protocoles expérimentaux (SAMVIQ, ACR, Pair Comparison, etc.) [110, 155]. Un 1<sup>er</sup> ensemble de tests psychophysiques avait pour cadre la mise au point des modèles de l'attention visuelle 2D et 3D avec des tests oculométriques sur images 2D de différentes bases, des tests de préférence de l'impact du flou sur la perception de la profondeur, des tests oculométriques pour l'étude du biais de profondeur (3D) et du biais central (2D et 3D). Un 2<sup>nd</sup> ensemble de tests [109, 137, 139, 177, 178, 179, 180, 182] avait pour objet l'évaluation de la qualité subjective de contenus décodés et/ou synthétisés avec, des tests de qualité sur images et vidéos visant à évaluer différents algorithmes d'interpolation de vues FVV (voir la figure 4.3), des tests de qualité d'images codées/décodées, des tests de qualité sur images 3D pour mesurer l'impact des algorithmes d'interpolation (visualisation 3D), des tests de la qualité perçue afin d'évaluer l'impact du codage des cartes de profondeur.

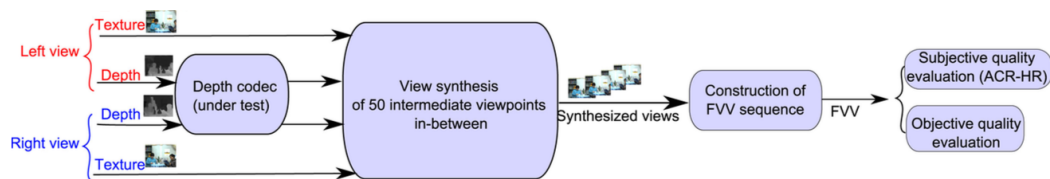


FIGURE 4.3 – Protocole expérimental pour l'évaluation des séquences FVV synthétisées (d'après [110]).

Ces résultats scientifiques issus de PERSEE ont été décrits et diffusés par le biais de livrables (20 rapports), de publications (10 articles de revues, plus de 35 communications à des conférences, de chapitres d'ouvrages collectifs [5, 108, 111, 126, 154]), de brevets et de propositions en normalisation internationale (7 soumissions [117, 149, 157, 167, 186, 231, 495] dont 2 définitivement acceptées). Quatre thèses de doctorat ont aussi été conduites et soutenues dans le cadre du projet [136, 145, 152, 171]. Les nouveaux outils et méthodes développés souvent collectivement, et visant un codage basé perceptuel des contenus 2D & 3D, ont été largement reconnus par la communauté scientifique. La plateforme logicielle intégrant les contributions des partenaires a aussi été développée. Enfin une base de données de vidéos stéréoscopiques servant de vérité de terrain a également été construite et diffusée [16].

## 4.4 Contributions et bilan personnels

J'ai participé au cycle de vie complet du projet PERSEE, depuis la 1<sup>re</sup> réunion entre partenaires en octobre 2008 pour la préparation de la soumission, jusqu'à l'audit final du projet en novembre 2013. Ma fonction de coordinateur du projet PERSEE, a évidemment amplifié l'ordre de mes contributions car j'ai dû remplir plusieurs fonctions. Le coordinateur a d'abord un rôle pivot au niveau de la communication interne et externe du projet. Pour accomplir cette tâche il doit veiller à ce que tous les outils nécessaires (site web, listes de diffusion, liste de contacts, serveurs d'échange, etc.) soient rapidement opérationnels. Il est l'intermédiaire entre l'ANR et les autres membres, et intervient pour :

- l'émission et le suivi des requêtes administratives.
- le suivi du calendrier de façon à anticiper les échéances et achever les livraisons, rapports, audits, etc.
- l'organisation des réunions de préparation entre partenaires.
- la production des documents (maquettes initiales, contrôles intermédiaires, finalisations).
- la maintenance et l'enrichissement des outils de communication.

Factuellement, après son lancement PERSEE s'est organisé autour de 19 réunions de travail entre partenaires, le projet a connu 2 audits (un à mi-parcours et le final), il a produit des rapports semestriels d'activités (et ceux finaux : justification des dépenses, résumé final public, etc.), et effectué 20 livraisons (essentiellement des rapports : états de l'art, documentations techniques sur les logiciels développés et la plateforme, descriptions et résultats des outils et méthodes).

Le coordinateur joue également le rôle d'animateur scientifique au sein du projet. Cette activité qui doit être modeste dans sa conduite, est néanmoins primordiale car le coordinateur doit veiller à ce qu'une liberté d'esprit et d'entreprise demeure au sein du projet, de façon à permettre à chacun d'exprimer et de développer sa recherche. Le coordinateur doit donc amener les hyper-spécialistes qui l'entourent, à entendre les problématiques des disciplines voisines, et en accepter les contraintes, les limites et les orientations. Par exemple dans le cas de PERSEE, il fallait faire comprendre que les tests de qualité subjective n'ont rien de génériques, et pour un nouveau service le test est conçu en tenant compte d'un très grand nombre de paramètres (conditions matérielles de visualisation, nombre et type d'observateurs, organisation des séquences de tests, nombre de vidéos à visionner, méthodes du test, recommandations à suivre, finalité des questions à poser aux observateurs, etc.).

Le détail de mes contributions d'ordre scientifique à la tâche 4 est dans le chapitre 8 (partie II).

L'objectif de PERSEE, concevoir un nouveau codage vidéo perceptuel basé sur l'amélioration de la qualité perçue, était un objectif ambitieux. Le projet a bien montré l'intérêt de l'approche en proposant tout un ensemble d'outils et de méthodes, où les choix pour améliorer la qualité perceptuelle guident le codage. Par rapport au projet ArchiPEG, PERSEE a impliqué 4 laboratoires et le projet se démarque par la masse et la diversités des résultats. La qualité de ces travaux a aussi été reconnue (somme de publications, brevets et contributions à la normalisation, 4 thèses soutenues). Cependant un verrou majeur demeure car il manque toujours un modèle de calcul de la distorsion perceptuelle pour la phase d'optimisation débit/qualité. Une métrique perceptuelle (et idéalement générique et facilement implantable dans le codeur), est indispensable pour concevoir un schéma de codage optimisé globalement en termes de qualité perçue. Une tel schéma permettrait de concurrencer les approches plus classiques qui sont encore prépondérantes au niveau des choix faits en normalisation. La recherche en codage vidéo continue donc de s'adresser à ces problématiques, via de nouvelles thèses et de nouveaux projets collaboratifs, et nous verrons au chapitre 5 que le projet PROVISION est dans la lignée des travaux de PERSEE.



## Chapitre 5

# Le projet ITN PROVISION

### 5.1 Contexte et objectifs

Le projet répond à l'appel « 2013 PEOPLE Programme (Marie Curie Actions) » du FP7. Le FP7 « 7th Framework Programme » est le septième programme cadre financé par l'Union Européenne (UE ) pour soutenir et développer la recherche et le développement technologique en son sein. Le FP7 fut déterminé pour la période 2007-2013 et il s'est poursuivi par le programme nommé « Horizon 2020 ». Le FP7 a exactement lancé quatre programmes spécifiques chacun répondant à des objectifs différents, parmi eux il y avait le programme « PEOPLE » visant à développer et renforcer le potentiel humain dans la recherche européenne. Ce programme PEOPLE prolonge les « actions Marie Curie » apparues avec le FP6 (on nomme à présent ces actions MSCA pour « actions Marie Skłodowska-Curie »), elles offrent des supports à la formation, la mobilité et le développement de la carrière de chercheurs européens. Les MSCA opèrent principalement soit en distribuant des bourses individuelles à de jeunes chercheurs, soit au travers le financement de réseaux ITN « Initial Training Networks » que constituent des institutions pour offrir en commun des programmes de formation à de jeunes chercheurs, les MSCA soutiennent aussi la coopération entre l'industrie et le milieu universitaire.

Fin 2012<sup>1</sup>, les partenaires font toujours le constat du rôle clé de la compression vidéo accompagnant les progrès des infrastructures réseaux pour le développement des services immersifs. Les progrès passent par des représentations encore plus efficaces des signaux vidéo tenant compte de leur dimension perceptuelle. Pour répondre à ces besoins, l'appel aux actions MSCA du FP7 offre l'opportunité de construire un consortium de rang international réunissant des institutions, académiques ou industrielles, complémentaires et leaders dans des domaines liés au codage vidéo. L'objectif commun est aussi de constituer un réseau de formation interdisciplinaire ayant pour objectifs de préparer

---

1. Exactement les partenaires ont déjà un an auparavant répondu sans succès à l'appel à projets FP7-PEOPLE-2012-ITN. La nouvelle tentative sera couronnée de succès, il faut noter que les conseils du service « Europe et recherche » de L'UNAM pour le montage du dossier ont été déterminants.

des chercheurs qualifiés, de trouver et valider des solutions efficaces et de normaliser les plus prometteuses.

Le projet scientifique du projet PROVISION<sup>2</sup> « PeRceptually Optimised Video CompressSION » ne s’inscrit pas simplement dans l’amélioration incrémentale du modèle de codage hybride classique (c’est-à-dire de tous les standards jusqu’à HEVC) intégrant les étapes de prédiction puis celles de transformation/quantification/codage de l’erreur de prédiction. Il adhère au modèle de codage vidéo basé contenu [159, 210, 251, 293, 323, 327, 356, 463] (on dit aussi basé synthèse, voir la figure 5.1) où l’approche vise la bonne qualité subjective du signal décodé sans chercher à minimiser directement les erreurs « pixel-à-pixel » entre les images originales et reconstruites. Il est ainsi possible de concevoir un système de compression où un cadre d’analyse/synthèse de textures remplace l’approche classique de minimisation de l’EQM pour le bénéfice de débits sensiblement inférieurs. Les approches novatrices intéressant le projet sont donc liées aux modules du modèle de la figure 5.1 avec la segmentation en objets [280, 393, 601], la synthèse de textures statiques et dynamiques [452, 459, 516], les métriques et l’évaluation de la qualité [161, 241, 270, 435], les techniques du codage résiduel [268, 290, 307, 399], l’optimisation débit-qualité [191].

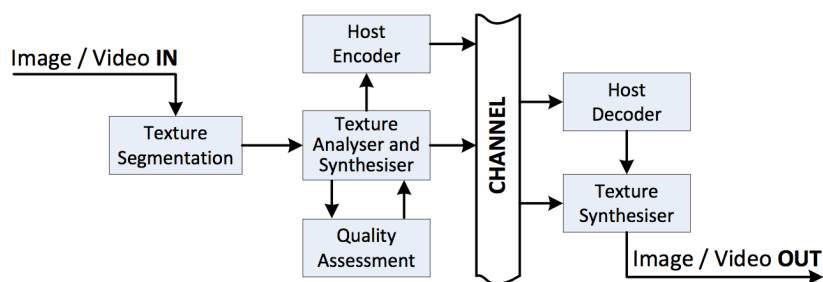


FIGURE 5.1 – Le modèle de codage vidéo basé contenu du projet PROVISION (d’après [151]).

## 5.2 Organisation

Le projet PROVISION a débuté en septembre 2013 pour une durée de 4 ans (fin en août 2017). L’ITN ambitionne alors de proposer le meilleur environnement de formation possible à un groupe de 15 jeunes chercheurs (12 doctorants nommés aussi ESR « Early-Stage Researchers », et 3 postdocs ou ER « Experienced Researchers ») afin de les conduire au plus haut niveau dans le domaine de la compression vidéo optimisée perceptuellement. Le consortium est constitué de cinq partenaires principaux qui bénéficient des fonds de l’UE pour recruter les ESR et ER et de partenaires associés qui les accueillent pendant leurs périodes de détachement. Les 5 partenaires principaux sont :

2. Le site web du projet PROVISION : [www.provision-itn.eu](http://www.provision-itn.eu)

**HHI** Fraunhofer HHI (son département « Image processing » à Berlin<sup>3</sup>), le HHI coordonne aussi le projet.

**UoB** Université de Bristol (« Bristol Vision Institute »<sup>4</sup>).

**BBC** British Broadcasting Corporation (sa division recherche & développement à Londres<sup>5</sup>).

**RWTH** Université de Aachen RWTH (« Institute of Communications Engineering » à Aix-la-Chapelle<sup>6</sup>).

**UoN** Université de Nantes (c'est-à-dire l'IRCCyN et son équipe IVC).

Il y avait 4 partenaires associés à l'origine du projet (Microsoft Research Asia, Purdue University, Technicolor et VITEC), pour des raisons variées (changement de priorités, difficultés administratives, etc.), ils furent finalement 3 (Queen Mary University of London, Technicolor et VITEC) à suivre le projet. L'aide allouée par l'UE est de 4 millions d'euros (dont 1,1 millions pour le partenaire UoN), 59% de l'aide va aux salaires des ESR et ER, et 21% aux frais liés directement à leur formation (organisation et participation à des activités).

Le programme de formation de l'ITN est complet et s'organise sur plusieurs niveaux. Il y a d'abord la formation principale (et individualisée) que l'ESR reçoit afin d'acquérir les compétences nécessaires pour faire avancer son projet de thèse et aller vers ses objectifs de carrière. Les activités sont alors supervisées par deux partenaires principaux (celui qui l'a recruté et où l'ESR conduit sa thèse, et un autre où le jeune fera 4 mois de stage), et un partenaire associé (où l'ESR sera détaché deux ou trois mois). Cette formation principale adresse donc un ensemble large de compétences (en recherche, technique, gestion de projets, communications, etc.). L'ITN propose en plus un programme de formations complémentaires dispensées lors de phases de rassemblement (une tous les 6 mois et organisée en alternance chez l'un des partenaires principaux). Ces formations complémentaires visent à développer, à travers différents types d'activités (cours, séminaires, travaux pratiques, travaux de groupe, etc.) des compétences ciblées (en management, communications, entrepreneuriat, sciences et techniques, etc.), et à renforcer l'esprit de groupe.

PROVISION est organisé en 7 lots de travaux (ou WP « Work Packages »), où les WP1 à 5 ont un caractère scientifique et interdépendant (voir aussi la figure 5.2) :

**WP1** Cas d'usages, besoins des services et comportement des utilisateurs : Définition des contextes pour les tests des technologies développées par les WP2 à 5.

**WP2** Modèles de mouvement avancés : Recherche de modèles paramétriques compacts pour la description des champs de vecteurs mouvement. Descriptions des objets en

---

3. Le site web du partenaire HHI : [www.hhi.fraunhofer.de/abteilungen/image-processing](http://www.hhi.fraunhofer.de/abteilungen/image-processing)

4. Le site web du partenaire UoB : [www.bris.ac.uk/vi-lab](http://www.bris.ac.uk/vi-lab)

5. Le site web du partenaire BBC : [www.bbc.co.uk/rd](http://www.bbc.co.uk/rd)

6. Le site web du partenaire RWTH : [www.ient.rwth-aachen.de/cms/research](http://www.ient.rwth-aachen.de/cms/research)

mouvement (considération de longues trajectoires, de leurs ruptures, etc.). Choix des images de références pour le codage vidéo.

**WP3** Méthodes de reconstruction 2D&3D : Développement de méthodes graduables d'analyse/synthèse de contenus pour le codage de scènes 2D, 2D+t et 3D.

**WP4** Outils perceptuels et évaluation de la qualité : Recherche de modèles perceptuels pour le traitement et le codage des vidéos 2D&3D. Développement d'outils pour la mesure de la distorsion perçue (métriques de qualité), et d'approches pour l'évaluation subjective de la qualité d'expérience.

**WP5** Modèles des tests et expérimentations : Développement de plateformes logicielles communes intégrant les outils et méthodes.

**WP6** Dissémination : Mise en place des outils de communication et de collaboration, et dissémination des résultats (en particulier contributions à la standardisation).

**WP7** Coordination et formation : Coordination et gestion du projet et de son programme de formation.

Les WP1 à 5 scientifiques sont chacun conduits par un partenaire particulier, mais les contributions peuvent être croisées. UoN mène évidemment le WP4 et s'intéresse à solutionner les verrous suivants (qui sont les sujets de 4 thèses financées) : la compression des textures dynamiques avec critère de fidélité perceptuelle (cadre de la thèse de Karam Naser), la conception de métriques de qualité à court terme, celle avec critère de fidélité sémantique, et la livraison de bout en bout de contenus avec contrôle de la qualité.

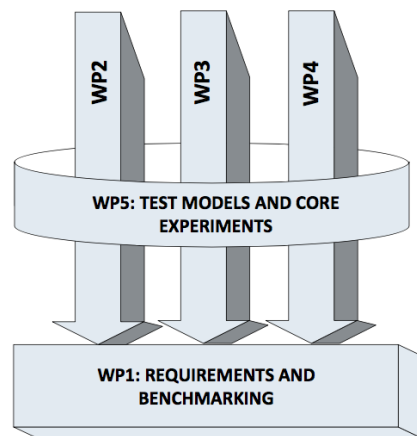


FIGURE 5.2 – Interdépendance des WP1 à 5 (c'est-à-dire des lots de travaux à caractère scientifique) de PROVISION (d'après [150]).

Le projet est piloté via plusieurs comités (voir la figure 5.3). Le « Supervisory Board », où chaque partenaire principal est représenté, contrôle et synchronise les activités scientifiques et de formation. L'« Advisory Board », où sont notamment présents les partenaires

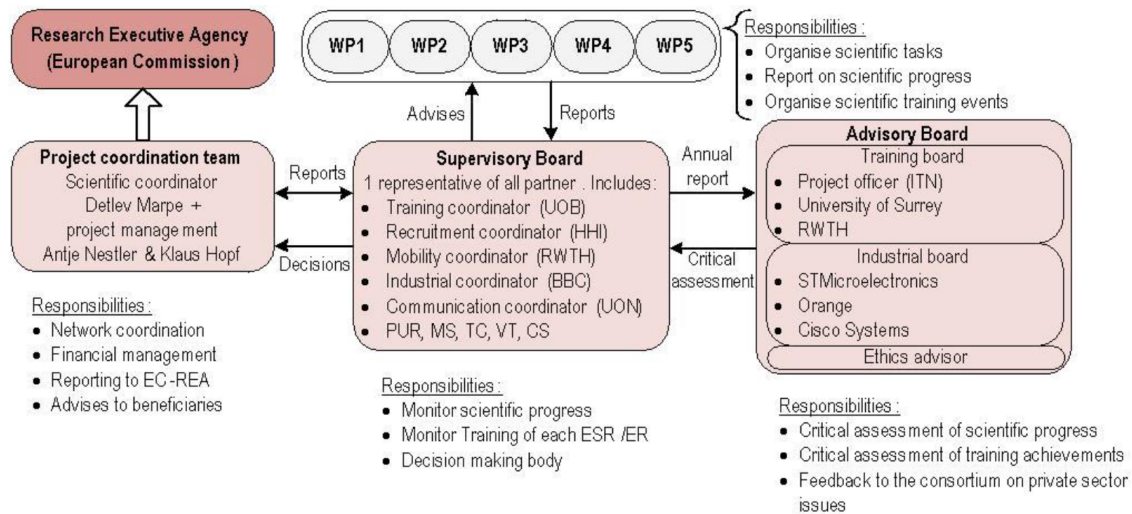


FIGURE 5.3 – Le pilotage de PROVISION (d’après [151]).

associés, évalue les avancées du projet. Le lien avec l’UE (aspects financiers, rapports d’activité, etc.) est réalisé via une équipe de coordination.

### 5.3 Contributions et bilan personnel

Mes contributions d’ordre organisationnel à PROVISION ont surtout été à partir de mi-2014 lorsque mes responsabilités à la tête du Département Informatique de Polytech’Nantes ont été moins prenantes, et jusqu’à la fin du projet en septembre 2017. Le projet se caractérisant par son organisation en réseau de formation, un programme complet d’activités a été proposé aux ESR&ER. Au niveau de l’équipe IVC qui a accueilli 3 doctorants (le 4<sup>e</sup> débutera sa thèse fin 2015), j’ai été chargé de coordonner leurs « outreach activities » qui visaient à promouvoir le projet en sachant communiquer auprès de publics différents. Par exemple nous participions à des évènements comme la « Fête de la science » (sensibilisation de scolaires à nos activités), la « Nuit européenne des chercheurs » (vulgarisation scientifique auprès d’un large public), au « Web2Day » et à « StartWest » (échange avec des acteurs du numérique locaux). J’ai aussi produit le modèle et rédigé le 1<sup>er</sup> ATAR [90] « Annual Training Activity Report » de PROVISION. Je suis également intervenu au niveau des démarches relatives à l’aspect éthique du projet, car son « Advisory Board » doit intégrer un conseiller éthique (voir la figure 5.3), et une conformité éthique (des procédures de tests avec observateurs) est demandée.

Par rapport aux deux projets ANR décrits précédemment, PROVISION a été remarquable de part sa dimension internationale et le prestige de ses partenaires qui ont tous eu un impact majeur sur les derniers standards liés au codage vidéo (avec des contributions aux consortiums ITU-T, ISO/IEC, 3GPP, DVB, IETF, VQEG, MPEG, W3C,

EBU), des personnalités du HHI et du RWTH président le développement des derniers standards de compression (présidence ou co-présidence du JVT et des groupes JCT-VC et JCT-3V). Les partenaires associés et/ou du comité « Advisory Board » (VITEC, Technicolor, Netflix, YouTube) ont aussi une expertise et une reconnaissance de niveau mondial. Le projet a donc offert l'opportunité de rencontrer directement et régulièrement ces acteurs, pour travailler au plus niveau dans le domaine du codage vidéo.

Deuxième partie

Contributions aux travaux de  
thèses

---

## Introduction

Cette partie présente mes contributions scientifiques au codage vidéo perceptuel. Les travaux ont été réalisés au travers l'encadrement de thèses, et également de certains masters de recherche ou de mémoires CNAM (dont la liste complète est à la section 1.5.1 du chapitre 1 exposant mon CV) qui se sont déroulés en amont des thèses (travaux de préparation), ou en parallèle (travaux d'accompagnement).

Je ne détaillerai pas ici les 1<sup>res</sup> recherches que j'ai conduites entre 2002 et 2004 à mon arrivée dans l'équipe IVC, c'est-à-dire celles menées dans le cadre de la thèse de Pierre Verbert, et pour laquelle nous avons déterminé une stratégie de codage efficace des projections issues de la transformation Mojette d'une image [36], ni ma participation au projet RNTL « Domus Videum », pour lequel nous avons proposé une indexation de plans vidéo, via une approche comparant des mosaïques 1D construites par quantification vectorielle algébrique et arborescente [35]. Ces travaux qui m'ont permis de renouer avec la recherche en codage d'images, n'ont cependant pas d'aspect perceptuel.

Les travaux qui sont présentés sont relatifs à la recherche d'outils et de méthodes pour améliorer la qualité perceptuelle du codage d'images et de vidéos 2D&3D. Nous avons agi au niveau de la couche de codage (définie dans l'introduction générale du chapitre 2) et développé des approches et outils comme, la transformation en ondelettes avec redressement orienté (à voir au chapitre 6), un module complet d'analyse des plans vidéo apte à piloter optimalement leur encodage (à voir au chapitre 7), des modèles d'attention visuelle 2D et 3D capables de guider le codage et de nouveaux services vidéo (à voir au chapitre 8), et un modèle de similarité visuelle entre textures exploitable par le codeur (au chapitre 9). Le dernier chapitre 10 présente les travaux de recherche en cours sur un codage de nuages de points 3D basé sur la quantification vectorielle.



## Chapitre 6

# Représentation d'images par transformées locales orientées et classes structurelles

### 6.1 Problématique

Fin 2004, nous nous intéressons avec Guillaume Jeannic et Dominique Barba, aux nouvelles représentations en ondelettes implantées par des techniques de redressement (ou « lifting »). Nous travaillons dans un contexte de compression d'images fixes, en effet le dernier standard est JPEG 2000<sup>1</sup> qui, en plus de performances en codage supérieures à JPEG, offre de multiples caractéristiques nouvelles (scalabilité, régions d'intérêt, résistance aux erreurs de transmission, codage sans perte). La nouveauté du point de vue compression, est que cette norme utilise une transformée en ondelettes (TO), qui possède une scalabilité en résolution naturelle, et permet le fonctionnement d'un algorithme de codage entropique sophistiqué basé sur l'algorithme EBCOT [519]. L'autre intérêt que nous y voyons, est que cette représentation plus riche de l'information image (car localisée en espace-fréquence, en résolution, et en orientation) a un potentiel certain pour un codage perceptuel, avec la possibilité d'adapter les lois de quantification aux sous-bandes de la TO en exploitant les propriétés du SVH (certaines étant déjà bien connues comme le phénomène de masquage visuel, ou la sensibilité au contraste [258]).

Cependant la TO 2D a des limites : simple extension bidimensionnelle de la TO 1D, elle n'est pas adaptée pour représenter efficacement des caractéristiques géométriques non horizontales ou non verticales. Le filtrage utilisé, isotrope suivant les 2 directions horizontales et verticales, ne peut s'adapter aux caractéristiques géométriques orientées d'une image[42] (contours, textures orientées, etc.). Pour pallier à ces limites, toute une nouvelle génération de transformées directionnelles a été développée (on les appelle

---

1. JPEG 2000 ou norme ISO/CEI 15444-1 commune à l'ISO, la DEI et l'UIT- T et produite par le groupe de travail JPEG « Joint Photographic Experts Group », voir aussi [166].

aussi les « X-lettres ») qui s'appuie principalement sur la TO. Certaines proposent une famille génératrice de fonctions orientées plus complète que les fonctions de base de la transformée en ondelettes (avec les « Ridgelets » [525, 537], les « Curvelets » [490], les « Contourlets » [382]). D'autres s'intéressent à une représentation adaptative de l'image. Soit l'image originale est déformée pour l'adapter au filtrage fixe de la transformée en ondelettes 2D [283], soit la base de la transformée est déformée pour s'adapter au contenu de l'image (avec les « Bandelets » [391, 400, 481], les « Curved Wavelets » [370], les « Directionlets » [368], l'« Adaptive Directional Lifting » [313], une approche de lifting avec échantillonnage quinconce [342], la « Direction-Adaptive Discrete Wavelet Transform » [310], etc.). Dans le cas des représentations adaptatives, les paramètres de déformation de l'image, ou de la base d'ondelettes doivent être transmis avec les coefficients de la transformée appliquée.

Partant des transformées adaptatives existantes, l'objectif de notre travail a été de définir une représentation qui tienne compte des orientations des caractéristiques géométriques de l'image. Une approche habituelle pour prendre en compte cette anisotropie passe par une modification des filtres utilisés pour chaque orientation de filtrage. Nous avons cherché dans notre approche à prendre en compte cette différence de traitement au niveau de la quantification en fonction de la statistique des sous-bandes produites par notre représentation, et de leur impact visuel.

Notre schéma de codage est illustré à la figure 6.1, on retrouve d'abord l'étape de transformation avec une représentation en ondelettes implantée par une technique de redressement. Ce dernier dépend de directions de régularité locale estimées via des approches basées extraction de contours ou minimisation d'une fonction de coût. Cette estimation de la géométrie de l'image produit une classification de l'activité directionnelle des blocs de l'image. La représentation des caractéristiques géométriques de l'image doit être codée pour être transmise. Une quantification adaptée des coefficients d'ondelette est aussi proposée en tenant compte à la fois des différentes classes structurelles des blocs et de l'anisotropie de la représentation.

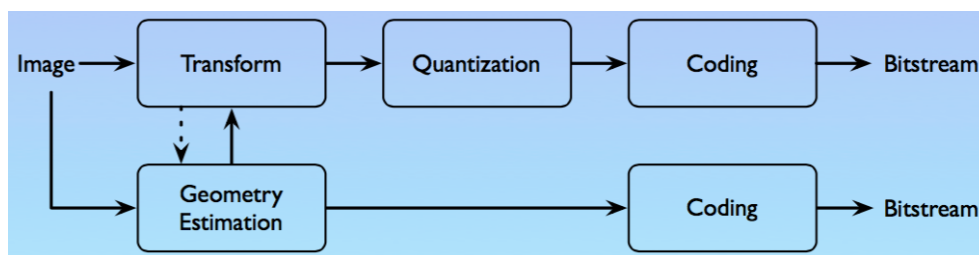


FIGURE 6.1 – Le schéma de codage utilisé (d'après la présentation de soutenance de thèse [278] disponible sur HAL).

## 6.2 Travaux menés et résultats

### 6.2.1 Etude de techniques de lifting orienté

La TO 1D peut être implantée par des techniques de lifting [539]. Ces techniques se décomposent en trois parties (voir la figure 6.2). Une décomposition polyphase qui sépare en deux ensembles les échantillons notés pairs et les échantillons notés impairs. Ensuite une succession d'étapes de prédiction et de mise à jour qui modifient l'un de ces ensembles par rapport à l'autre, et vice versa. Et enfin une étape de mise à l'échelle de ces éléments. La TO 2D standard est alors obtenue en appliquant successivement une TO 1D implantée par lifting selon les lignes puis selon les colonnes (ou inversement).

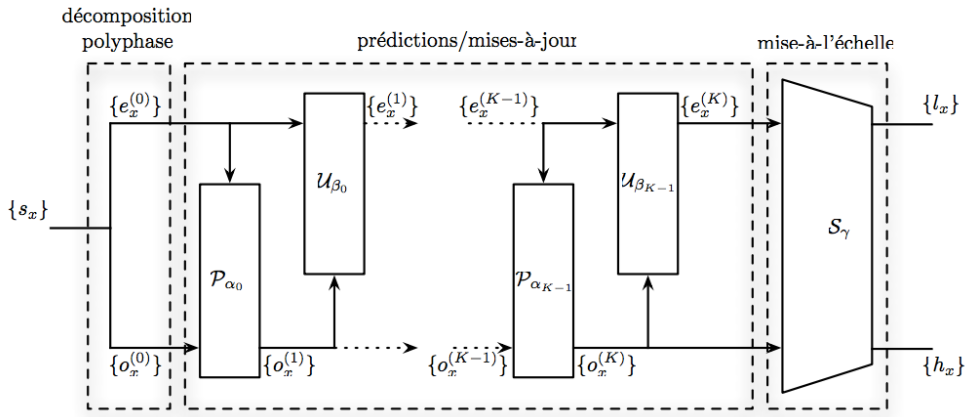


FIGURE 6.2 – Les étapes de la technique de lifting pour un signal 1D (d'après [278]).

La flexibilité de ces techniques de lifting a permis de développer de nouvelles TO 2D orientées qui s'adaptent au contenu de l'image à représenter [370]. Contrairement à la TO 2D standard qui n'opère le filtrage que selon deux directions fixées horizontale et verticale, ces transformées orientées procèdent selon deux directions adaptatives. Pour cela, le principe général de ces techniques réside dans les choix des échantillons, ceux pris dans l'ensemble utilisé pour prédire, et ceux pris dans l'ensemble utilisé pour mettre à jour. Exactement les étapes de prédiction et de mise à jour sont obtenues en considérant des échantillons qui ne sont pas sur la même ligne pour un filtrage d'orientation plutôt horizontale, ou sur la même colonne pour un filtrage d'orientation plutôt verticale (voir la figure 6.3). Pour implanter une telle transformée, une solution consiste à prédire les éléments des colonnes impairs (respectivement lignes) à partir d'une combinaison linéaire des éléments des colonnes paires (respectivement lignes) et de mettre ensuite à jour les éléments paires à partir des erreurs de prédiction précédemment calculées. Comme la transformée en ondelettes 2D standard, la reconstruction reste exacte et la décimation critique.

À partir de cette définition, différentes implantations pratiques sont possibles. Parmi celles-ci trois techniques de lifting orienté ont été étudiées. La 1<sup>re</sup> dite « 8-connexe-

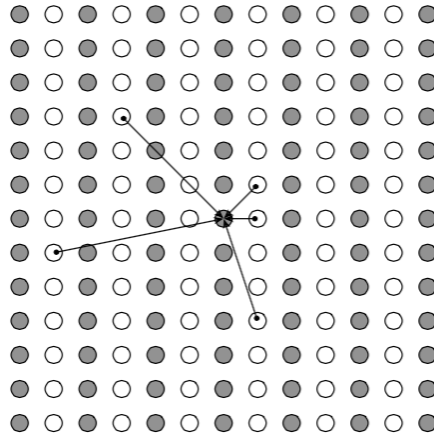


FIGURE 6.3 – Exemple, d'un schéma lifting basé colonne, avec la prédiction d'un échantillon d'une colonne impaire (en gris) à partir de divers échantillons de colonnes paires (en blanc) (d'après [278]).

(p,q)-lifting » [41, 368] est illustrée à la figure 6.4(a) où le schéma de redressement est appliqué selon des droites discrètes 8-connexes orientées selon un vecteur (p,q). La 2<sup>e</sup> dite « (p,q)-connexe-(p,q)-lifting » [310, 368] est illustrée à la figure 6.4(b), où l'orientation de filtrage est toujours définie par un vecteur (p,q), mais les droites discrètes sont (p,q)-connexes au lieu d'être 8-connexes (cela revient à une TO sur une grille discrète ré-échantillonnée). Enfin la 3<sup>e</sup> dite «  $\theta$ -lifting » [313, 41] et illustrée à la figure 6.4(c), dans ce cas  $\theta$  définit l'orientation de filtrage, les valeurs utilisées pour la prédiction et la mise à jour sont interpolées à partir d'échantillons de parité appropriée. L'interpolation est choisie linéaire de façon à assurer une reconstruction exacte.

Nous avons effectué une étude comparative des trois implantations du redressement orienté [39, 278] en les appliquant sur différents types de contenus (synthétiques ou naturels, textures ou contours, structures rectilignes ou courbes) et pour une gamme complète d'orientations. La comparaison directe est impossible car dans le cas du  $\theta$ -lifting, la dynamique du signal de la sous-bande des hautes fréquences varie selon l'orientation de filtrage. La solution a été de comparer la capacité des trois implantations à minimiser l'énergie de l'image reconstruite uniquement à partir de la sous-bande des hautes fréquences le long de l'orientation de filtrage. Les expérimentations sur images synthétiques (non bruitées) ont montré l'avantage de la technique « (p,q)-connexe-(p,q)-lifting » pour les orientations  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  et  $135^\circ$ . Sur images naturelles de contenu mono-orienté, l'effet lisseur de l'interpolation permet d'améliorer légèrement les performances du  $\theta$ -lifting. La méthode basée sur des droites discrètes 8-connexes n'a quant-à-elle jamais donné satisfaction.

Finalement nous proposons une représentation basée sur une décomposition en ondelettes selon deux directions qui s'adaptent localement, la 1<sup>re</sup> orientation de filtrage correspondant à l'orientation dominante localement [23]. Des filtres différents peuvent être

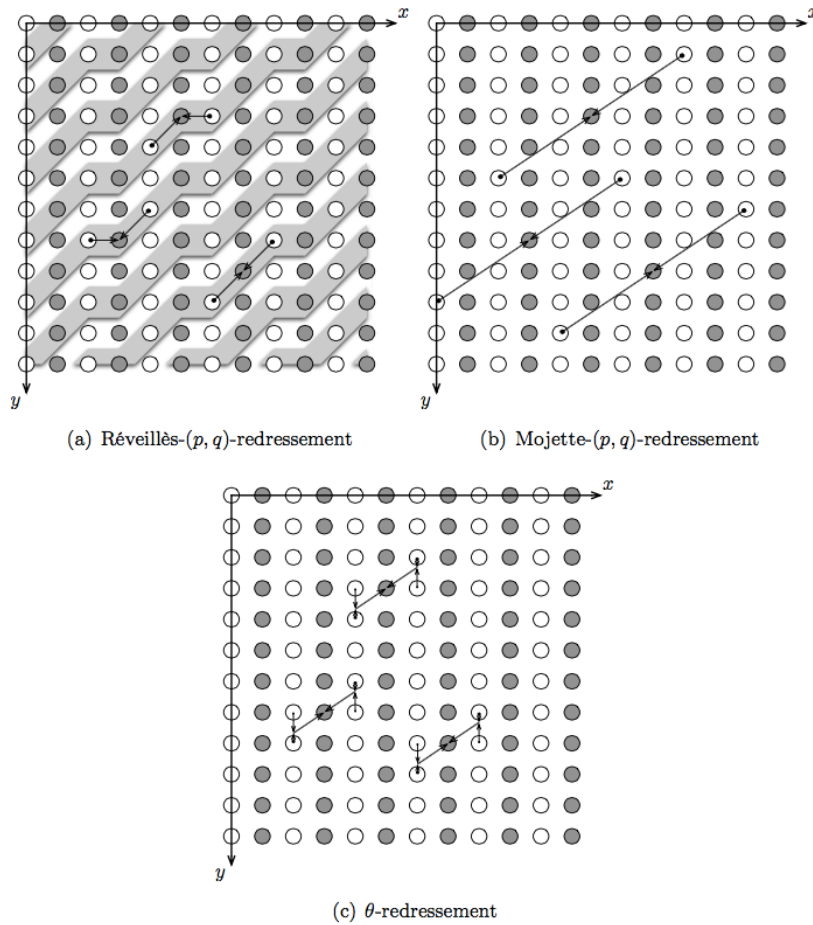


FIGURE 6.4 – Pour les 3 implantations testées, les exemples de l'étape de prédiction du schéma lifting basé colonne, pour une orientation  $(p, q)$  de filtrage d'environ  $33^\circ$  par rapport à l'horizontal (d'après [278]).

appliqués selon chacune des deux orientations de filtrage, permettant ainsi un traitement anisotrope. Comme la TO 2D classique, notre représentation produit 4 sous-bandes pour un niveau de décomposition donné :

**BF** Les basses fréquences selon les deux directions de filtrage.

**HFReg** Les hautes fréquences selon la direction de régularité, et les basses fréquences selon la 2<sup>nd</sup>e direction de filtrage.

**HFOrtho** Les hautes fréquences selon la 2<sup>nd</sup>e direction de filtrage, et les basses fréquences selon la direction de régularité.

**HFRes** Les hautes fréquences selon les deux directions de filtrage.

La décomposition polyphase de notre TOO (transformations en ondelettes orientées) varie d'un bloc de l'image à l'autre, car elle dépend du choix de la 1<sup>re</sup> orientation de filtrage (voir aussi la figure 6.5). Ainsi (idéalement) les structures considérées dominantes seront uniquement présentes dans la sous-bande HFOrtho alors que la sous-bande HFReg ne contiendra que le bruit le long de ces structures.

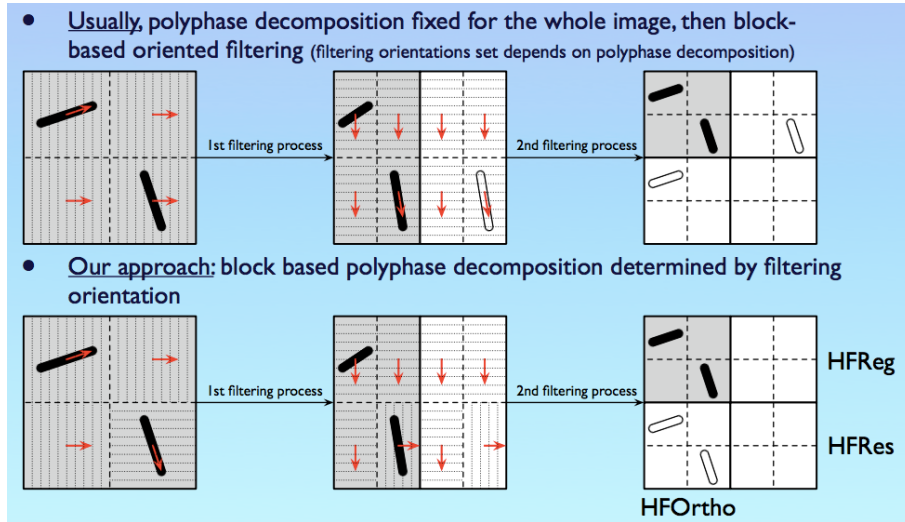


FIGURE 6.5 – Comparaison de l'implantation de notre TOO par rapport à une plus classique (d'après la présentation de soutenance de thèse [278] disponible sur HAL).

## 6.2.2 Estimation de la géométrie de l'image

La TOO ne s'applique donc pas selon une seule et unique orientation de filtrage sur toute l'image. Son intérêt réside dans son adaptation au contenu de l'image, et l'orientation de filtrage est définie localement. La segmentation de l'image en régions de même orientation est généralement considérée par blocs. Le choix des orientations de filtrage n'est pas trivial avec la méthode utilisant l'interpolation (c'est-à-dire avec le  $\theta$ -lifting) car l'énergie des filtres équivalents aux diverses orientations de filtrage possibles varie en fonction de l'orientation. Deux méthodes ont été alors proposées pour déterminer l'orientation principale des blocs de l'image.

Une 1<sup>re</sup> méthode [23] est basée sur une extraction des contours significatifs de l'image. Après chaînage et estimation de l'orientation des éléments de ces contours, les régions de l'image découpée en blocs sont classifiées, et pour chaque bloc une orientation de filtrage correspondant (si elle existe) à une orientation dominante de la région est assignée (voir la figure 6.6). La 2<sup>de</sup> méthode [41] est basée sur la mesure de l'orientation du gradient, en modélisant l'incertitude de cette mesure en tout point par une loi gaussienne. Ces mesures pour chaque bloc sont alors cumulées et le nombre de maxima ainsi que l'écart avec une loi uniforme sont utilisés pour classifier le bloc. Ces deux méthodes fournissent à

la fois une classification des régions de l'image découpée en blocs et un choix d'orientation de filtrage pour ceux-ci, ces deux informations servant de base à notre TOO.

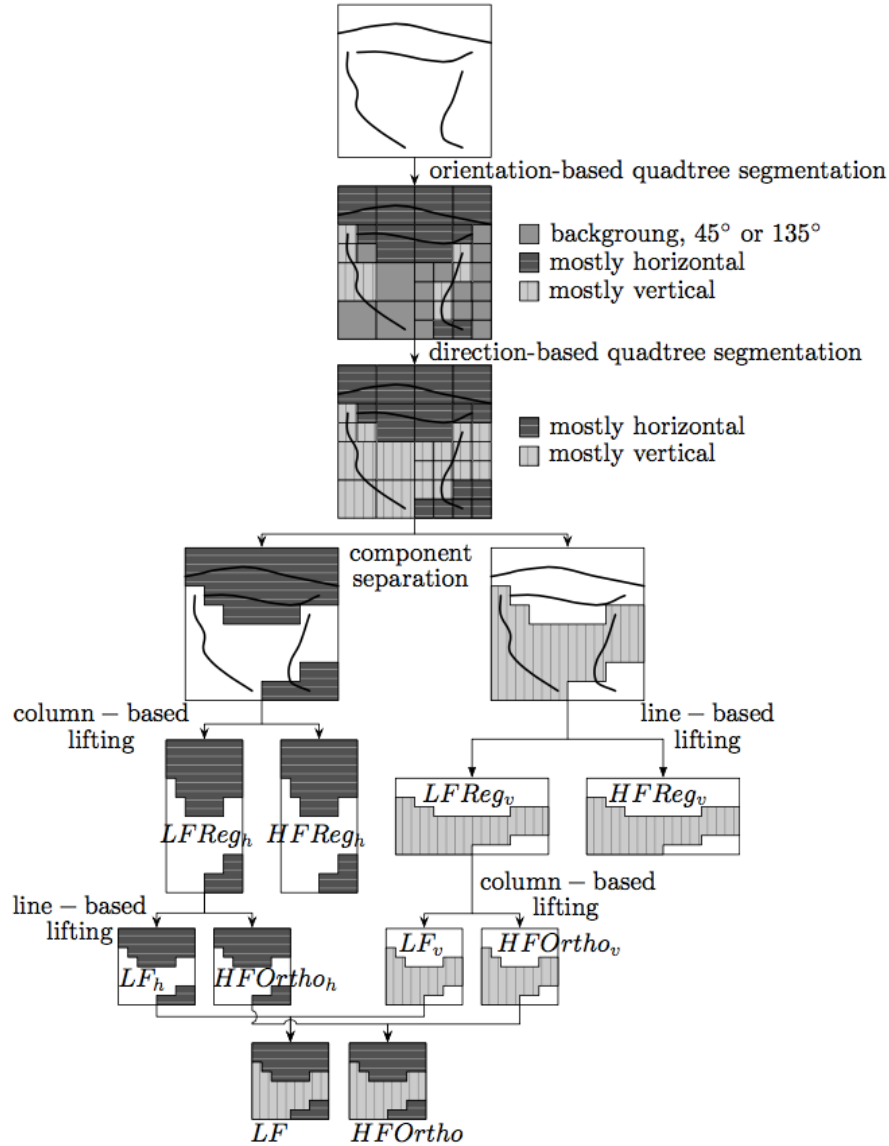


FIGURE 6.6 – Les étapes de notre TOO (d'après [23]).

Ces deux méthodes (celle [a] basée extraction de contours, celle [b] basée mesure du gradient) définissent 4 classes qui caractérisent l'activité des régions de l'image, cependant ces classes ne désignent pas exactement le même contenu selon la méthode employée (et donc ne conduisent pas exactement à la même segmentation). Pour chacune de ces classes, une orientation de filtrage est définie :

**zone uniforme** Selon la méthode [a] : zone ne contenant aucun élément de contour extrait. Selon [b] : zone où moins d'un certain quorum d'échantillons ont leur norme du gradient supérieure à un certain seuil. Pour la zone uniforme, aucune orientation de filtrage n'est préconisée, cette dernière est choisie arbitrairement (par exemple  $0^\circ$  ou  $90^\circ$  pour correspondre à une TO conventionnelle).

**zone (ou texture) mono-orientée** Selon [a] : zone où le pourcentage d'éléments orientés de contours est supérieur à un seuil haut. Selon [b] : zone où la présence d'une orientation privilégiée unique n'est pas due au bruit. Pour la zone mono-orientée, l'orientation de filtrage choisie est celle dominante de la région considérée, afin de tirer parti de la régularité le long de cette direction et minimiser localement l'énergie de la sous-bande des hautes fréquences produite. Cette direction de filtrage est nommée direction de régularité.

**zone (ou texture) multi-orientée** Selon [a] : zone où le pourcentage d'éléments orientés de contours est supérieur à un seuil bas. Selon [b] : zone où la présence d'au moins deux orientations privilégiées n'est pas due au bruit. De même pour la zone multi-orientée, l'orientation de filtrage choisie est celle dominante de la région considérée. Cette fois la sous-bande des hautes fréquences après le premier filtrage, contiendra les structures qui correspondent aux autres orientations de la région.

**zone texturée de façon isotrope** Selon [a] : zone où le pourcentage d'éléments orientés de contours n'est pas supérieur à un seuil bas. Selon [b] : zone où aucune orientation privilégiée n'est détectée. Pour la texture isotrope aucune orientation de filtrage n'est évidente, on la choisit arbitrairement comme dans le cas de la zone uniforme.

### 6.2.3 Codage de la géométrie de l'image

Les deux méthodes [a] (basée extraction de contours) et [b] (basée mesure du gradient) du paragraphe précédent 6.2.2 aboutissent donc chacune à deux cartes : l'une relative aux orientations de filtrage, l'autre aux classes structurelles des régions de l'image. Chacune des deux cartes peut être représentée via un arbre quaternaire. Seule la méthode [a] construit les arbres à partir d'un résultat intermédiaire : les contours extraits de l'image. Pour que cette dernière soit reconstruite au décodeur, l'information supplémentaire (c'est-à-dire les contours pour la méthode [a], ou bien les 2 cartes pour la méthode [b]) doit aussi être transmise.

#### Codage arithmétique adaptatif des contours de l'image en fonction de la résolution inférieure

Dans le contexte de codage de contours extraits à différentes résolutions d'une image (voir la figure 6.7), une méthode pour coder les chaînes d'éléments de contours à un niveau de résolution donné, et en utilisant l'information disponible aux résolutions



inférieures, a été proposée [22]. En effet l'écart entre la direction orthogonale à l'orientation du gradient à la résolution inférieure, et l'orientation d'un morceau du contour déjà codé/décodé à la résolution courante, est corrélé avec la position du prochain élément du contour à coder/décoder (voir la figure 6.8).

En pratique l'implantation par un codeur arithmétique adaptatif de notre méthode surpasse les méthodes de l'état de l'art utilisant un modèle de chaîne de Markov d'ordre 1 à 3. Toutefois la combinaison de notre méthode en collaboration avec un modèle de Markov n'a pas été concluante pour des problèmes de dilution de contexte. Malgré cela la diminution mesurée du débit avec notre méthode par rapport à un modèle de chaîne de Markov d'ordre 1, est en moyenne de 3% sur une moyenne de dix images et atteint jusqu'à 8%. Notre méthode est toutefois moins efficace que la méthode de l'état de l'art dans certains cas où une structure extraite à la résolution courante n'est pas présente à la résolution inférieure. Pour ces cas de figure, il pourrait être intéressant de mettre en compétition les deux méthodes au lieu de les utiliser en combinaison.

### **Codage arithmétique adaptatif des arbres quaternaires définissant les orientations de filtrage et les classes de structure**

Ces deux arbres partagent la même structure. Une implantation de leur codage a été proposée, elle s'appuie sur un codage arithmétique adaptatif basé sur une exploitation de la cohérence spatiale des nœuds de l'arbre précédemment codés/décodés. La structure de l'arbre précisant si un nœud est terminal (feuille) ou non, est codée en fonction de la profondeur, c'est-à-dire la probabilité qu'un nœud soit une feuille est liée à sa profondeur, cette probabilité est initialisée par une loi uniforme et s'adapte au fur et à mesure du codage/décodage. Les valeurs des feuilles de l'arbre sont aussi codées arithmétiquement, avec un contexte de passé d'ordre 1. Les valeurs d'orientation de filtrage et de classe d'activité ne sont pas codées indépendamment : nous tenons compte du fait que les régions de l'image classifiées « uniforme » ou « texture isotrope » ont une orientation de filtrage arbitrairement fixée. Le contexte de passé est l'information des feuilles déjà codées/décodées jouxtant la feuille à coder/décoder. À cette fin, une carte des classes et des orientations de filtrage déjà codées/décodées est mise à jour au fur et à mesure du codage/décodage.

Le coût de codage des éléments de contours extraits d'une image, a été comparé au coût de codage des structures en arbre calculées à partir de ces mêmes contours. L'information est préalablement quantifiée plus ou moins finement, via les paramètres que sont les seuils utilisés pour la classification des blocs de l'image, et via la taille minimale possible de ces blocs qui détermine aussi la profondeur maximale des arbres. L'expérimentation a montré que, même pour une taille minimale de blocs de 4x4 pixels, le coût de codage des chaînes de contours restent beaucoup plus élevé que le codage des arbres quaternaires associés [278]. Ceci est dû en partie à la redondance liée aux contours présentant une même orientation dans une même région de l'image.

#### 6.2.4 Quantification perceptuelle des coefficients d'ondelettes orientées

L'objectif est de tirer parti des estimations des orientations de filtrage et de classification structurelle pour adapter le procédé de quantification des coefficients des différentes sous-bandes produites par notre TOO. Nous voulons aussi exploiter, et plus efficacement que JPEG 2000 ne le fait, la propriété de masquage visuel du SVH selon laquelle le seuil de détection d'un stimulus n'est pas fixe, mais lié à la valeur moyenne locale du signal d'image l'environnant. On parle alors de signal masquant versus signal masqué. Dans notre cas le masquant est l'image originale, et celui à masquer est constitué des erreurs introduites par la quantification. Il faut donc être capable d'estimer correctement les effets de masquage au sein de l'image afin de, par exemple, déterminer le pas de quantification le plus grand qui n'introduira pas de changement perceptible dans l'image décodée.

Plusieurs solutions existent dans JPEG 2000 pour prendre en compte cet effet de masquage visuel [166]. Une fait appel à un modèle simple où le signal masquant est considéré comme n'appartenant qu'à un domaine fréquentiel restreint en un site donné sans tenir compte des autres signaux environnants. Le potentiel de masquage d'un tel signal dépend à la fois de sa fréquence et de son amplitude. Les fonctions CSF « Contrast Sensitivity Function » sont alors utilisées, chacune indiquant pour une distance donnée de visualisation, la sensibilité aux variations de contrastes en fonction de la fréquence spatiale mesurée en cycles par degré (voir la figure 6.9). En pratique, il faut pondérer le pas de quantification de chaque sous-bande (donc d'une gamme de fréquence donnée) du signal transformé par une valeur CSF (des tables de pondérations sont disponibles dans la littérature).

Une autre solution utilise le fait que l'effet de masquage d'une décomposition en ondelettes est une fonction monotone croissante (monotone si on ne considère pas l'effet de facilitation) de l'amplitude du coefficient d'ondelettes [492]. Donc plus l'amplitude d'un coefficient d'ondelettes est élevée, plus dans son entourage local l'effet de masquage est important, et plus le bruit de quantification peut y être élevé sans introduire d'artefact visuel. Cependant cette modélisation du masquage n'est valable que si un coefficient d'ondelettes représente un seul et unique signal en terme de fréquences spatiales. Or la TO 2D classique est telle que ce n'est pas nécessairement le cas comme le montre la figure 6.10. Le pire exemple y est illustré avec la sous-bande « HH » de l'espace spatio-fréquentiel, elle contient les détails diagonaux correspondent aux signaux de l'image orientés autour de  $-45^\circ$  (ou  $135^\circ$ ) et  $45^\circ$ . Or pour le SVH, les signaux orientés à  $45^\circ$  ne peuvent pas masquer ceux orientés selon  $-45^\circ$ , et inversement. Une superposition de tels signaux antagonistes entraînant une valeur de coefficient d'ondelettes élevée sera considérée à tort comme masquante par le modèle. L'effet masquant sera surestimé, la quantification résultante sera trop grossière, et les distorsions seront finalement visibles. Notre approche de TOO doit pouvoir surmonter cette limite.

Expérimentalement nous procédons en comparant la TO classique 2D sur 5 niveaux de décomposition avec notre TOO sur le même nombre niveaux de décomposition. Pour la TOO, les orientations de filtrage sont déterminées à pleine résolution pour une taille

de bloc de 16x16 pixels, et les orientations retenues sont utilisées pour les 3 premiers niveaux de décomposition. Pour les deux derniers niveaux où il n'y a pas de structures exploitables, une transformation en ondelettes classiques 2D est appliquée. Pour les 3 premiers niveaux de décomposition une stratégie est adoptée pour la quantification en tenant compte des différentes classes structurelles :

**Zones uniformes** L'information visuelle est contenue dans la sous-bande basse fréquence BF. Pour exploiter cette propriété nous éliminons les coefficients localisés dans les sous-bandes haute fréquence en les mettant à zéro.

**Zones mono-orientées** L'information des contours est principalement localisée dans la sous-bande HFOrtho, le pas de quantification pour cette dernière doit être relativement petit en comparaison des pas utilisés respectivement pour les sous-bandes HFReg et HFRes.

**Zones multi-orientées** L'information structurelle est présente dans plus d'une sous-bande, aucune ne peut donc être privilégiée par rapport aux autres.

**Zones texturées isotropes** Si l'activité est grande, un facteur de masquage visuel peut être appliqué [492], et une quantification grossière n'introduira pas d'artefact visuel.

Nous avons développé une méthodologie expérimentale via une application ad'hoc (voir la figure 6.11) permettant de régler aux différents niveaux de résolution, les pas de quantification des sous-bandes et les pondérations des classes structurelles, tout en observant les variations de qualité perçue des images codées/décodées. L'application procure le moyen de vérifier si l'approche fait baisser le coût de codage de l'image tout en maintenant sa qualité une fois décodée, la qualité est aussi calculée objectivement via différentes métriques (C4 [270], SSIM [435], PSNR). Les améliorations sont notables dans les zones uniformes grâce à la quantification plus fine permise par la réduction du débit dans les autres régions sans perte de qualité, et aussi dans les régions filtrées selon des orientations diagonales. Cependant la procédure est complexe, les résultats variables dépendent de la nature structurelle des images, et l'absence d'une métrique objective de qualité totalement fiable interdit un réglage automatique du système (par exemple via une optimisation de type débit-distorsion).

### 6.3 Discussion

Les limites de la TO 2D pour la représentation et le codage des images fixes étaient connues. De nouvelles représentations pour remédier à ces limites avaient été proposées, et nous avons choisi d'en considérer une famille : les TO orientées basées sur des techniques de redressement. Les objectifs de notre travail étaient de définir une telle représentation, et de concevoir, en tenant compte de caractéristiques locales variables, une quantification adaptative basée sur des critères perceptuels, puis son codage. Si nous

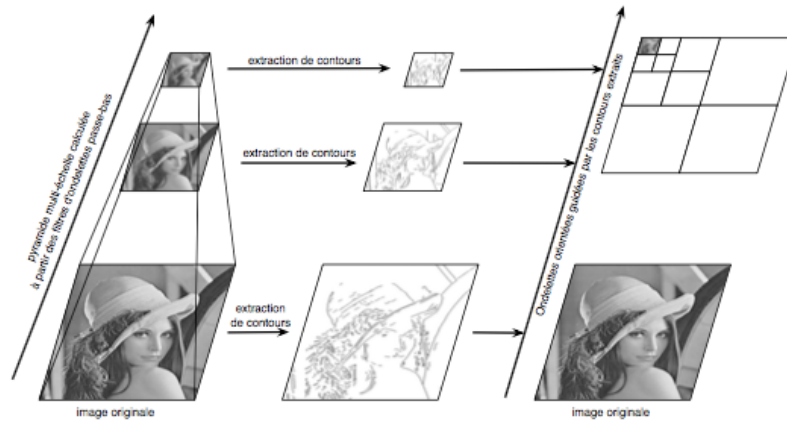
avons pu démontrer l'intérêt de l'approche, faute d'une métrique perceptuelle générique fonctionnelle, nous n'avons pas réussi à produire un schéma de codage compétitif.

Depuis les travaux ont continué pour l'adaptation des transformées, le plus souvent directement motivés par la normalisation menée en compression vidéo. De nombreuses recherches ont donc concerné la DCT « Discrete Cosine Transform », outil incontournable intégré aussi bien à H.264 qu'à HEVC. Par exemple avec la DCT orientée [225, 295] offrant plus de mode de directions, l'ajout de balayages optimisés [271] et l'amélioration des matrices de quantification [222]. On peut noter à ce niveau le regain d'intérêt pour les transformations avec chevauchement limitant les effets de blocs [552], certaines ayant été implantées dans H.264/AVC [184, 273], d'autres retenues par des schémas concurrents (par exemple [450] utilisé par le projet Daala<sup>2</sup>). D'autres approches ont visé à construire de nouvelles bases de fonctions directionnelles par apprentissage, en particulier la MDDT « Mode Dependent Directional Transform » où l'apprentissage dépend des modes de prédiction [294] (ce procédé a été intégré au KTA puis simplifié pour HEVC [174, 190, 201, 208, 209, 218]). Des approches d'apprentissage plus robustes ont été proposées [202, 261], certaines optimisant directement le critère débit-distorsion [213, 244], ou d'autres critères pour classifier [173], et s'adapter au contenu [214]. Pendant la standardisation d'HEVC, de nouvelles approches ont été avancées (mais pas incluses au standard) pour transformer les blocs résiduels intra ou inter, par exemple le schéma AMT « Adaptive Multiple Transform » constitué d'un ensemble de transformées de la famille DCT et DST « Discrete Sine Transform ». Il y a eu l'introduction du concept de transformations secondaires [175] exploitant les corrélations qui subsistent au sein des coefficients de la 1<sup>re</sup> transformation. Dans la même famille, il y a eu cette structure fonctionnant en deux temps [142], le 1<sup>er</sup> utilisant une transformée de Haar 2D visant à concentrer l'énergie résiduelle des basses fréquences, et le 2<sup>nd</sup> mettant en œuvre une transformation non-séparable 2D suivant 4 directions (horizontale, verticale, 45° et 135°) pour exploiter les corrélations subsistantes. Une autre approche récente de transformation adaptée au contenu [104] opère d'abord un sous-échantillonnage du bloc résiduel en 4 sous-blocs, le 1<sup>er</sup> étant utilisé pour construire la base des transformations adaptées aux 3 autres blocs.

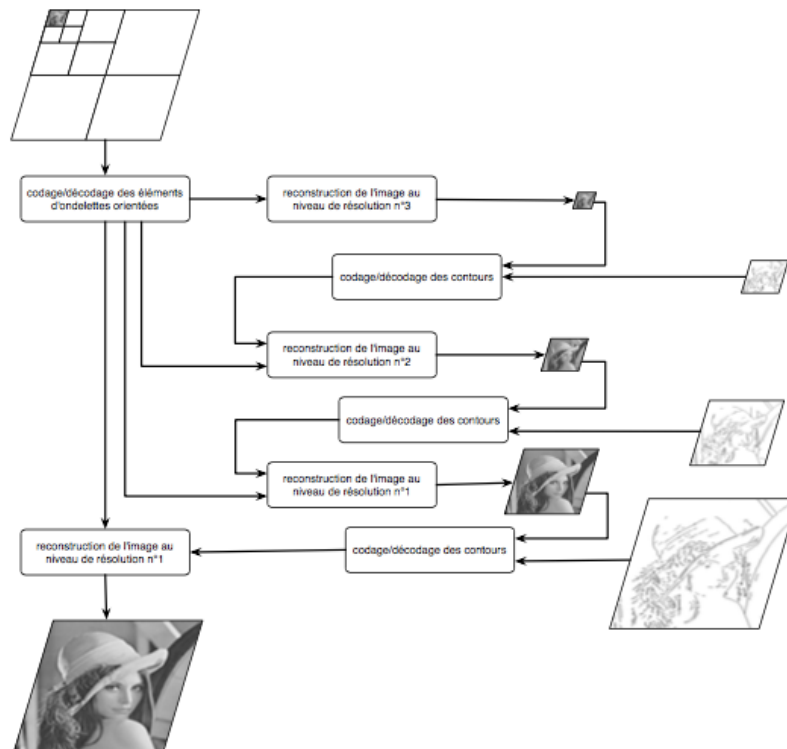
La recherche de transformations adaptées demeure manifestement un champ d'investigation toujours très actuel. En compression vidéo, la nature du signal à transformer, des erreurs résiduelles après prédiction intra ou inter, est bien différente du cas des images fixes avec JPEG 2000. Elle va donc dépendre des performances des prédicteurs. Si des gains de codage non négligeables ont été obtenus via des transformations adaptatives pour H.264/AVC, l'énergie résiduelle est plus réduite avec HEVC [122] et donc plus complexe à exploiter. L'amélioration a été apportée en définissant des TU « Transforms Unit » de taille variable, et en employant plusieurs transformations.

---

2. Le site web du projet Daala : [xiph.org/daala](http://xiph.org/daala)



(a) Extraction des contours à chaque niveau de résolution obtenu par une transformée en ondelettes 2D, suivi d'une transformée en ondelettes orientées.



(b) Codage et décodage des contours et des éléments d'ondelettes orientées.

FIGURE 6.7 – Schéma de codage/décodage des contours extraits utilisé dans un contexte de TOO (d'après [278]).

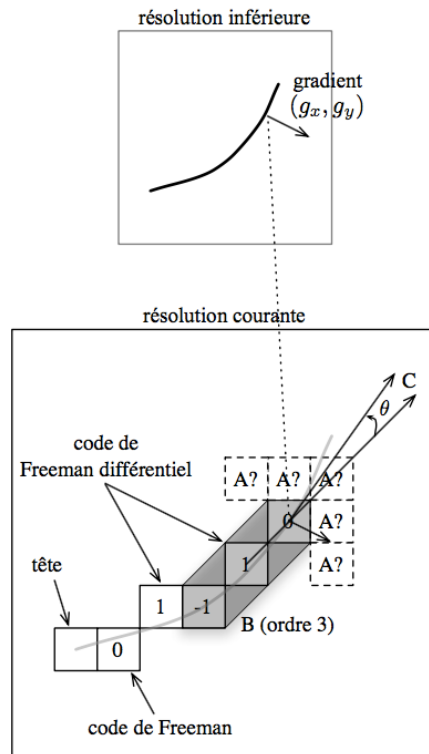


FIGURE 6.8 – Approche multi-résolution pour le codage arithmétique de contours. A : prochain code différentiel de Freeman à coder/décoder. B : les N derniers codes différentiels de Freeman codés/décodés. C : écart entre la direction orthogonale au gradient à la résolution inférieure et la direction donnée par les 2 derniers pixels codés/décodés. (d'après [22]).

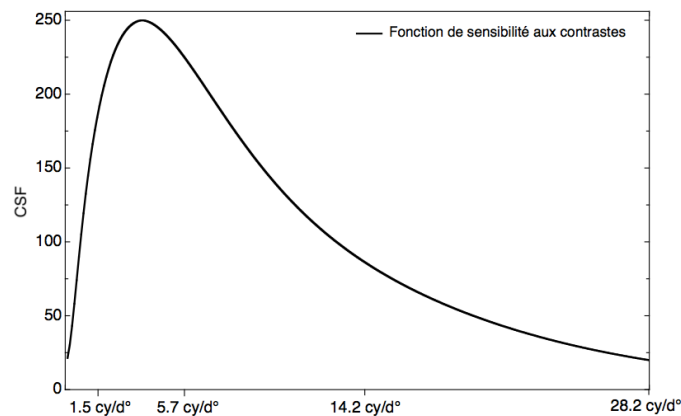


FIGURE 6.9 – Exemple d'une CSF montrant que l'œil humain est plus sensible aux variations d'amplitude d'un signal moyenne fréquence que haute fréquence (d'après [278]).

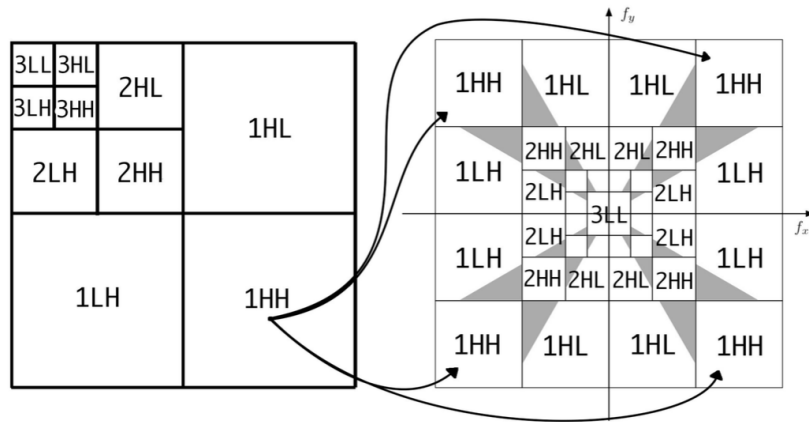


FIGURE 6.10 – À gauche : la décomposition en TO 2D en sous-bandes représentée dans un espace spatio-fréquentiel. À droite : la décomposition en ondelettes dans l'espace fréquentiel (d'après [278]).

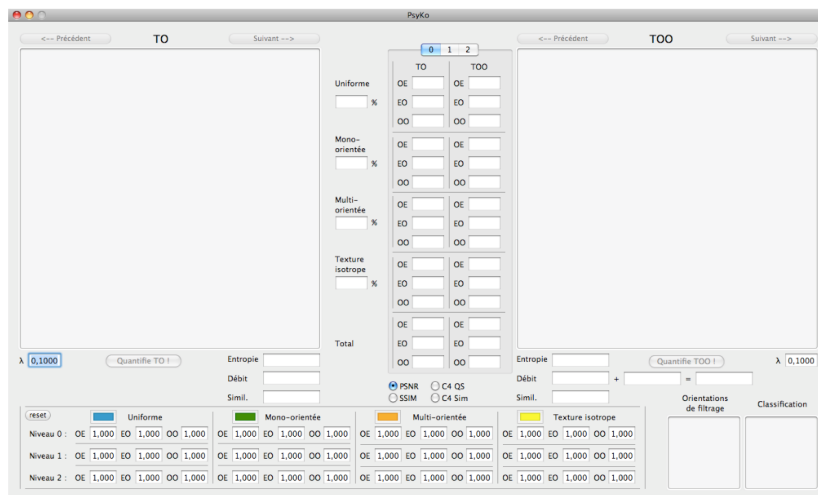


FIGURE 6.11 – Capture d'écran de l'application permettant de modifier les pas de quantification et les pondérations structurales, tout en observant les images codées/décodées (d'après [278]).

## Chapitre 7

# Outil de pré-analyse avant encodage vidéo perceptuel

### 7.1 Problématique

Début 2006, avec Olivier Brouard, Fabrice Delanay, et Dominique Barba, nous entamons le travail relatif au « Sous-projet 4 : Pré-analyse et conditionnement du flux vidéo HD » du projet ArchiPEG (voir le chapitre 3).

Le nouveau standard de codage vidéo H.264/AVC, combine de nombreuses techniques permettant de compresser les vidéos beaucoup plus efficacement que les normes précédentes, et fournit plus de flexibilité aux applications. Par exemple, le modèle de référence (appelé aussi modèle de vérification) que nous utilisons, implante (voir aussi la figure 7.1) : une compensation de mouvement effectuée à partir de 5 images de référence, mettant en œuvre 7 modes de prédiction inter (sur blocs  $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ ,  $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$  et  $4 \times 4$ ), et une précision au quart de pixel pour décrire les déplacements ; deux modes de prédiction intra (sur blocs  $16 \times 16$  et  $4 \times 4$ ) ; un filtre anti effet de bloc intégré ; le codage entropique CABAC ; etc.

Les travaux de recherche visent donc optimiser ce codeur complexe, et nous distinguons trois grandes familles d’approches. La 1<sup>re</sup> est relative aux méthodes d’estimation de mouvement et vise à accélérer la recherche du vecteur déplacement [299, 322, 334, 345, 354, 366, 371, 373, 381, 406, 410, 449, 461, 469, 475, 478, 479, 493, 520, 558, 565, 626, 628], il faut par rapport à la méthode de recherche exhaustive réduire la complexité calculatoire tout en maintenant un niveau suffisant de qualité objective pour la vidéo décodée. La 2<sup>e</sup> famille regroupe les techniques permettant de réduire les modes de prédiction (intra et/ou inter) testés [286, 317, 319, 320, 325, 328, 335, 337, 338, 340, 346, 349, 352, 353, 357, 359, 364, 369, 374, 375, 379, 384, 385, 396, 407, 411, 415, 437, 438, 460, 476]. Elles procèdent en constituant, en fonction des caractéristiques spatio-temporelles du macrobloc courant et de sa corrélation avec son voisinage, un ensemble plus réduit (car plus probable) de modes à tester. La 3<sup>e</sup> famille est constituée des techniques de



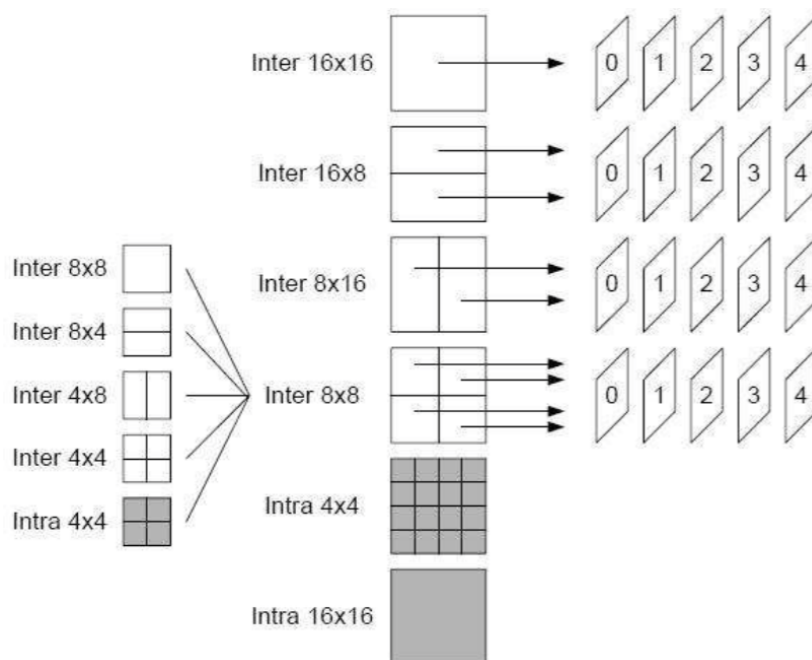


FIGURE 7.1 – Les étapes de la recherche exhaustive des prédictions intra et inter du logiciel de référence de H.264/AVC (extrait de [217]).

sélection des images de référence [318, 347, 350, 365, 416, 425, 433, 446, 458, 469]. Par défaut le codeur utilise pour l'estimation du mouvement, les 5 images précédentes de l'image courante, alors que les expériences ont montré la prépondérance en moyenne de l'image de référence la plus proche temporellement de celle courante. Les techniques visent donc à éviter les recherche parmi les multiples images de référence possibles.

Nous faisons donc le bilan que le codeur H.264/AVC ne possède pas les outils nécessaires pour prendre des décisions permettant d'assurer un codage cohérent fonction du contenu spatial des images et de leur évolution temporelle. Le codeur prend les décisions à court terme et seulement selon des considérations de type débit-distorsion où cette dernière est appréciée via une mesure de type EQM. Un artéfact typique des vidéos décodées, est l'effet de papillotement (« flickering effect ») apparaissant sur les textures à bas débit et dû aux changements successifs de modes de la prédiction et du pas de quantification des macroblobs.

Nous nous sommes aussi naturellement intéressés aux propriétés et modélisations du SVH, pouvant être utilisées pour améliorer les performances des codeurs d'images. Les limites des métriques de qualité traditionnelles (SAD, MSE, PSNR) sont connues [638], et de nouvelles, intégrant CSF et effets de masquage, ont été proposées [166, 467, 468, 492, 506, 523, 535, 557, 581, 590, 591]. Elles demeurent cependant encore trop complexes ou dédiées à des applications spécifiques. Quelques rares approches en codage vidéo mettent en avant des modèles perceptuels [305, 355, 435], mais la majorité concerne les aspects

allocation binaire et quantification [398, 543, 554, 583, 597], cependant aucune n'a été adoptée par les standards, où les efforts ont porté sur l'enrichissement des procédés de décorrélation spatiale et temporelle, le codage entropique, etc.

Nous avons regardé les propriétés du SVH, en tenant en compte plus particulièrement de la dimension temporelle primordiale pour la vidéo : perception de la luminance et des couleurs [584, 603, 606, 624, 629, 636, 648], sensibilité au contraste [253, 436, 441, 588, 611, 638, 647], organisation multi-canal [487, 498, 507, 536, 540, 557, 566, 615, 621, 633], et effets de masquage [405, 498, 551, 576, 579, 582, 585, 592, 610, 644, 646]. Ces propriétés ont un intérêt pour la conception de modèles de l'attention visuelle (AV). L'AV désigne le mécanisme de sélection des informations spatio-temporelles pertinentes du monde visible [100, 578, 587, 598, 643], elle repose sur deux processus interagissants [429, 436, 528, 529, 533, 553, 567, 602, 619, 630, 632] :

- un mécanisme pré-attentif (appelé aussi exogène ou « bottom-up »), où la sélection éphémère et involontaire est directement guidée par les informations saillantes du champ visuel.
- un mécanisme attentif (appelé aussi endogène ou « top-down »), où l'attention et le déplacement oculaire s'effectuent sous un contrôle volontaire et cognitif, guidés par une tâche à accomplir.

Modéliser entièrement l'AV est une tâche trop complexe, et notre cadre d'utilisation (c'est-à-dire la visualisation de la TVHD) est limité, nous nous sommes donc concentrés sur la modélisation de l'AV pré-attentive. Quelques approches pour construire ces modèles sont empiriques (c'est-à-dire basées sur des méthodes de traitement classique d'image) [513, 518, 544, 562, 568], quelques autres sont statistiques (c'est-à-dire reposent sur l'hypothèse que la capacité d'attraction d'une région est inversement proportionnelle à sa probabilité d'apparition) [444, 466], la plupart sont basées sur une architecture liée à la théorie de l'intégration de caractéristiques visuelles (ou FIT « Feature Integration Theory », voir aussi la figure 7.2) [613, 632]. Les nombreux modèles [237, 256, 257, 308, 351, 388, 436, 445, 448, 477, 496, 511, 512, 542, 587, 594] diffèrent alors par la richesse des caractéristiques visuelles considérées (typiquement : contraste, mouvement, couleur, orientation, etc.), et par la méthode de fusion des différentes cartes de caractérisation en une seule de saillance. Quelques travaux font référence à la dimension temporelle [237, 256, 388, 451, 487, 505, 543], et le mouvement a clairement été identifié comme un attracteur visuel [611]. Quelques exemples prometteurs d'applications de modèles d'AV au sein de codeurs vidéo existent. L'idée est de favoriser la qualité des zones plus importantes visuellement. On parle de compression sélective ou de compression avec régions d'intérêt. Deux méthodes (qui peuvent être complémentaires) permettent de l'effectuer : la 1<sup>re</sup> dite indirecte [388], consiste à altérer judicieusement le signal à coder de façon à réduire la quantité d'information sur les zones de moindre intérêt. La 2<sup>de</sup> dite directe [388, 443, 543], modifie le cœur du codage (l'allocation en débit) en fonction de la connaissance des régions d'intérêt.

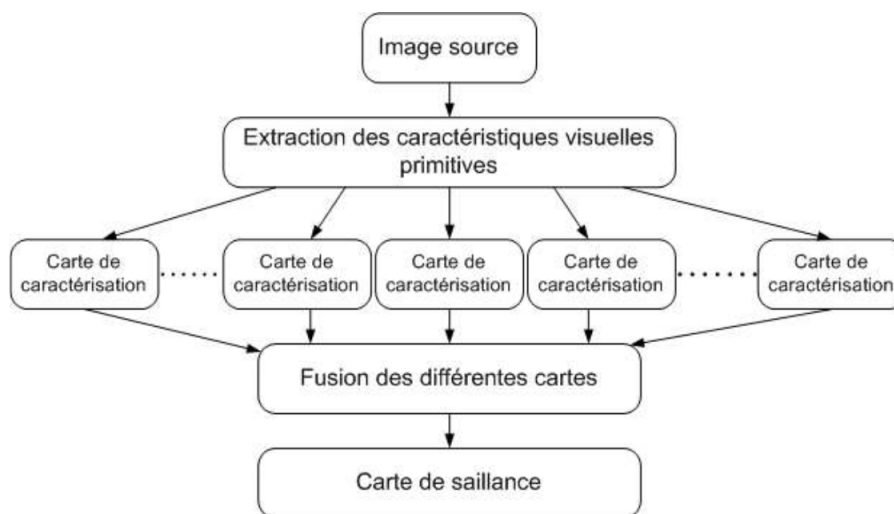


FIGURE 7.2 – Architecture biologiquement plausible d’un modèle d’attention visuelle (extrait de [613]).

Enfin, notre solution est de contraindre le codeur vidéo à ne plus prendre des décisions à court terme mais de le guider dans ses choix afin d’assurer la cohérence du codage d’un objet tout le long de sa durée de vie. Le codeur H.264/AVC ne possédant pas les outils nécessaires pour analyser le contenu vidéo, l’obtention des informations doit être réalisée par une étape de pré-analyse en amont du codeur (voir la figure 7.3). L’analyse porte sur un segment temporel homogène dont la taille est fonction du temps de fixation du SVH (environ 200 ms [351, 388], soit 9 « frames » successives de la TVHD en mode progressif). Le cœur de l’outil d’analyse est présenté à la figure 7.4. Le traitement intra segment vise à détecter les objets vidéo, le contraste de mouvement est le facteur déterminant utilisé. Le traitement inter segment permet de suivre les objets pendant leur cycle de vie. A l’issue de cette combinaison de traitements intra et inter segments, l’outil produit pour chaque segment temporel, une carte de segmentation et une description détaillée (forme, mouvement, couleur, texture, etc.) de chaque objet suivi. L’outil intègre ensuite un modèle d’AV qui combine les informations de contraste spatio-temporel, pour calculer la carte de saillance visuelle du segment. L’ensemble des informations (c’est-à-dire pour chaque segment, une carte de segmentation et une carte de saillance visuelle) est transmis au codeur afin de déterminer des paramètres de codage adaptés (par exemple avec H.264/AVC, pour structurer le GOP, assurer la cohérence des prédictions intra et inter des macroblocs appartenant à une même région, sélectionner des images de référence, quantifier en fonction de l’attraction visuelle). La suite détaille la construction de l’outil de pré-analyse avant codage vidéo.

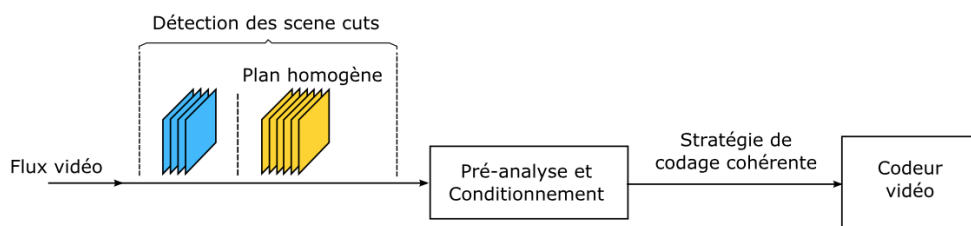


FIGURE 7.3 – Pré-analyse de la vidéo en amont du codeur, avant traitement la vidéo a été découpée en plans homogènes (extrait de [217]).

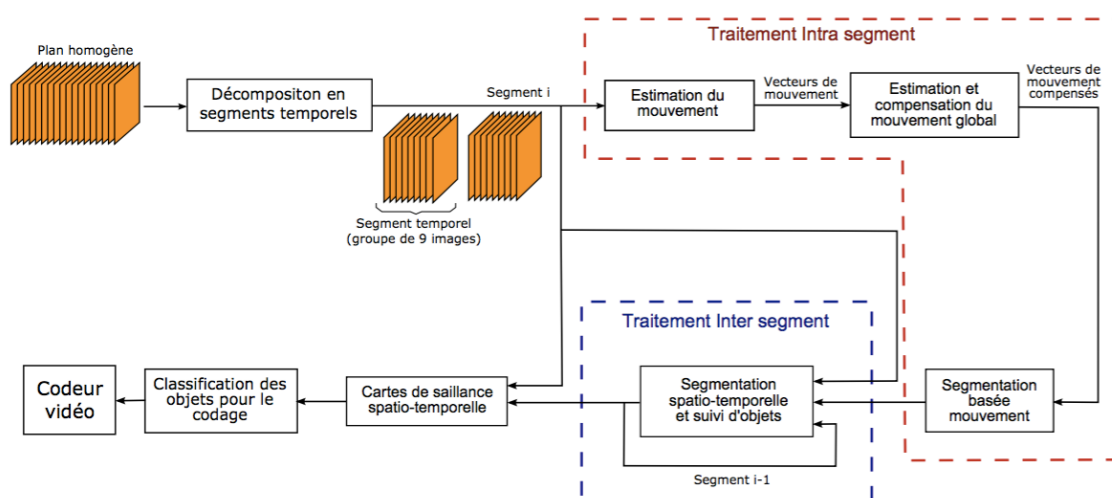


FIGURE 7.4 – Le cœur de l'outil de pré-analyse vidéo (extrait de [217]).

## 7.2 Travaux menés et résultats

### 7.2.1 Segmentation basée mouvement

Nous devons segmenter la vidéo en objets vidéo [431, 447, 484, 521, 625], dans notre cas un tel objet est une forme spatio-temporelle caractérisée par une couleur, une texture et un mouvement propre qui diffère du mouvement global du segment (c'est-à-dire souvent celui dû au capteur). Nous considérons cette dernière caractéristique comme prioritaire et réalisons d'abord une segmentation basée mouvement.

#### Estimation du mouvement basée tubes spatio-temporels

La 1<sup>re</sup> étape est l'estimation du mouvement au sein du segment temporel, pour cela nous introduisons la notion de tube spatio-temporel (ou tube de mouvement, voir la figure 7.5) qui suit et aligne un macrobloc<sup>1</sup> donné sur plusieurs images successives. Pour

1. Par la suite, on parlera indifféremment de bloc ou de macrobloc.

lisser les vecteurs déplacement et obtenir un champ homogène, nous faisons l'hypothèse d'un mouvement uniforme.

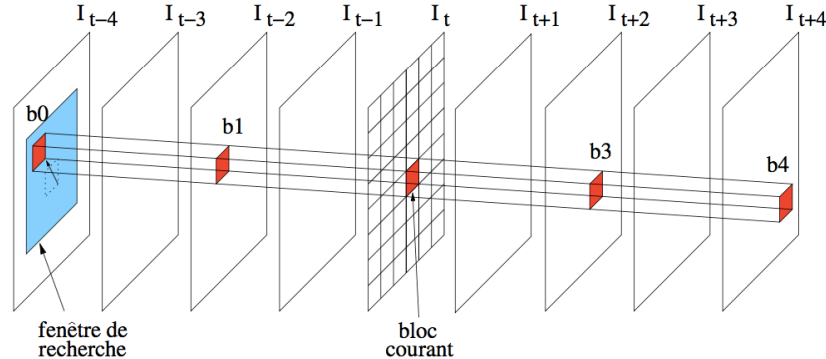


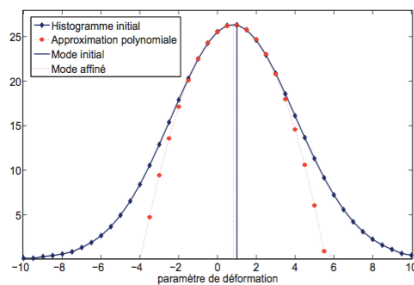
FIGURE 7.5 – Tube spatio-temporel (extrait de [40]).

Pour réduire la complexité (cf. contraintes temps réel du projet ArchiPEG), nous ne retenons que 5 images pour calculer les vecteurs déplacement : l'image courante située au centre, 2 images passées et 2 futures (la figure 7.5 l'illustre). Les images HD sont filtrées spatialement et sous-échantillonnées par un facteur 6 en 2 temps (d'abord un facteur 2 puis celui 3) [329]. Avant chaque sous-échantillonnage, un filtre passe-bas adéquat est appliqué. A partir des images filtrées et sous-échantillonnées, l'estimation du mouvement est effectuée. Chaque bloc de l'image courante est simultanément comparé aux blocs potentiellement correspondants des images précédentes et futures. L'erreur globale est obtenue par la somme de 4 erreurs quadratiques moyennes, chacune entre le bloc courant et ses 4 blocs correspondants passés et futurs (et en tenant compte des composantes YUV). Le vecteur de mouvement retenu est celui qui minimise l'erreur globale. L'estimation est multi-résolution car les vecteurs estimés à une plus faible résolution, sont multipliés par un facteur approprié afin d'être utilisés comme point de recherche initial à la résolution supérieure. La méthode produit un champ de vecteurs homogène reflétant les mouvements apparents dans la scène, exactement il y a un vecteur déplacement par tube, et un tube associé à chaque macrobloc de l'image centrale du segment temporel.

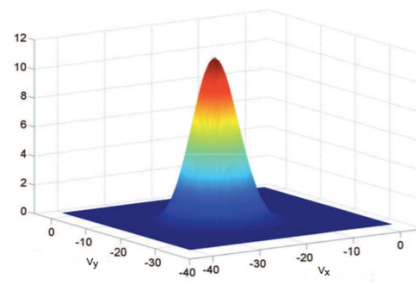
### Estimation robuste du mouvement global

La 2<sup>e</sup> étape est l'estimation des paramètres du modèle de mouvement global du segment temporel. Plusieurs modèles paramétriques existent [394, 574], nous retenons un modèle affine simple à 6 paramètres (2 paramètres caractérisent la translation, les 4 autres les déformations de rotation et de zoom). Pour estimer ces paramètres, nous adaptons une méthode [40, 417] qui repose sur l'accumulation au sein d'histogrammes des dérivées orientées (gradients) des vecteurs mouvement (exactement, à chaque gradient est associé un poids gaussien). Le mode majoritaire d'un histogramme représente la valeur du paramètre retenu, pour affiner la localisation du mode, une estimation des

moindres carrés est utilisée (voir la figure 7.6(a)). Dans un premier temps, les 4 paramètres de déformation sont calculés (un histogramme utilisé par paramètre), puis les vecteurs déplacement sont compensés à partir des 4 paramètres de déformation précédemment obtenus. La dernière étape est l'accumulation des vecteurs compensés dans un histogramme 2D afin d'extraire les paramètres de translation (voir la figure 7.6(b)). La méthode d'estimation du mouvement global est robustifiée car la contribution de chaque vecteur mouvement est pondérée en fonction d'un indice de confiance. Ce dernier dépend de l'activité spatiale (c'est-à-dire des gradients spatiaux) des macroblocs du tube de mouvement (voir la figure 7.7). Les contributions peu fiables (par exemple celles issues des blocs uniformes) sont donc écartées. Un exemple de compensation du mouvement global d'un segment temporel, est illustré à la figure 7.8 avec la séquence vidéo<sup>2</sup> *New mobile and Calendar*.



(a) Histogramme d'accumulation pour estimer un paramètre de déformation.



(b) Histogramme d'accumulation pour estimer un paramètre de translation.

FIGURE 7.6 – Estimation par accumulation des paramètres du mouvement global d'un segment temporel (extrait de [217]).

### Segmentation basée mouvement

Pour la 3<sup>e</sup> étape nous reconsidérons l'espace d'accumulation des vecteurs caractérisés par les 4 paramètres de déformation. Le mode principal correspond au mouvement global du segment temporel, les autres modes représentent les mouvements des autres objets, et nous réalisons donc leur segmentation dans cet espace [21] (voir la figure 7.9). La première étape consiste à éliminer le bruit : un seuil de rejet est défini empiriquement et toutes les cellules de l'histogramme inférieures à ce seuil sont mises à zéro. L'algorithme traite ensuite récursivement les modes par ordre décroissant à partir du principal. Pour chaque position connexe à un mode, un gradient orienté est calculé (c'est-à-dire la différence des populations entre celle du mode et celle de la cellule considérée). Tant que le gradient est positif, la position testée est considérée comme appartenant au mode, et l'opération répétée pour les cellules connexes. L'algorithme est réitéré tant qu'il reste des

2. Nous avons utilisé des séquences HD du SVT [491], elles sont disponibles là : [ftp.ldv.e-technik.tu-muenchen.de/pub/test\\_sequences](http://ftp.ldv.e-technik.tu-muenchen.de/pub/test_sequences)

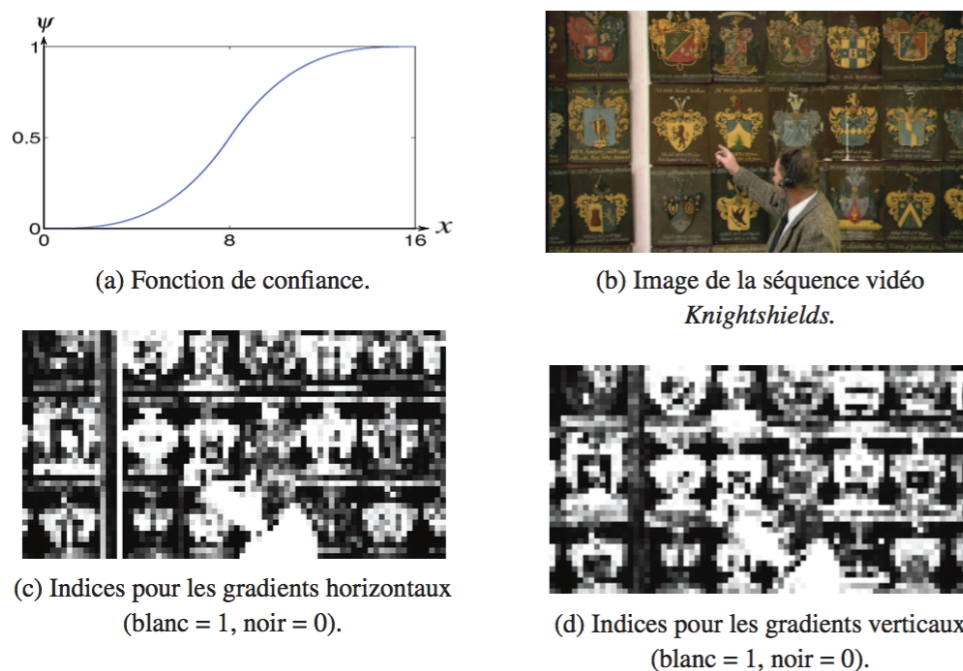


FIGURE 7.7 – Indices de confiance des vecteurs mouvement (extrait de [40]).

cellules non nulles n'appartenant à aucun mode. Les quelques cellules marquées comme appartenant à plusieurs modes, sont définitivement rattachées au plus proche.

### 7.2.2 Segmentation spatio-temporelle et suivi des objets

Le critère unique du mouvement n'est pas suffisant pour obtenir une segmentation en objets vidéo compacts, des caractéristiques supplémentaires sont considérées, 3 sont intra-segment (connexité spatiale, couleur, et texture), et une inter-segment (connexité temporelle). L'objectif est aussi d'assurer le suivi des objets d'un segment temporel à l'autre. Une approche markovienne est retenue [19, 348, 570] pour intégrer tous ces critères (voir la figure 7.10).

Nous exprimons les propriétés du champ markovien par une distribution explicite. Soit  $E = \{E_s, s \in S\}$  le champ des étiquettes des objets sur l'ensemble  $S$  des sites  $s$  (c'est-à-dire des tubes), les réalisations de  $E$  sont notées  $e = \{e_s, s \in S\}$ . Soit  $O = \{O_s, s \in S\}$  le champ des observations, leurs réalisations sont notées  $o = \{o_s, s \in S\}$ . Soit  $\Lambda$  (respectivement,  $\Omega$ ) l'ensemble des réalisations possibles de  $E$  (respectivement, des configurations d'étiquettes possibles de  $e$ ).  $\eta = \{\eta_s, s \in S\}$  est la structure de voisinage définie sur  $s$ .  $(E, O)$  est donc modélisé par un champ de Markov aléatoire. Le champ d'étiquettes optimal  $\hat{e}$  est obtenu selon un critère MAP (« Maximum A Posteriori »). Le théorème de Hammersley et Clifford établit l'équivalence entre champs markoviens et distributions de Gibbs [637], la configuration optimale de  $\hat{e}$  est alors obtenue en

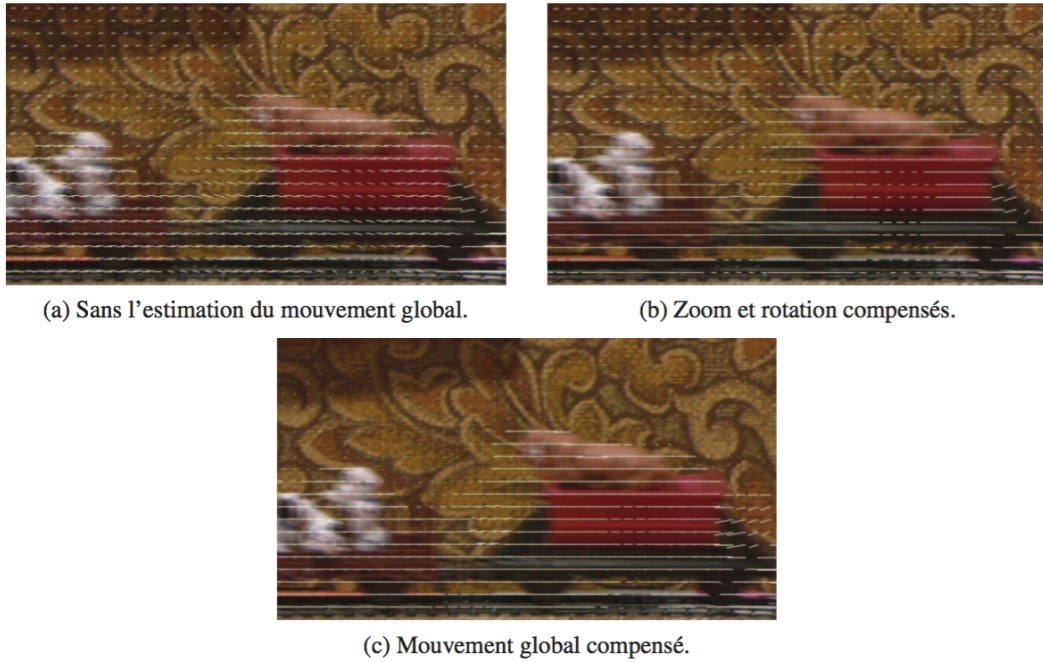


FIGURE 7.8 – Exemple de compensation du mouvement global de la séquence vidéo *New mobile and Calendar* (extrait de [40]).

minimisant une fonction d'énergie globale  $U(o, e)$  :

$$\hat{e} = \arg \min_{e \in \Omega} U(o, e)$$

Les propriétés markoviennes permettent d'écrire cette fonction d'énergie comme la somme de fonctions de potentiels élémentaires  $V_c$  définies localement sur des cliques [618] :

$$U(o, e) = \sum_{c \in C} V_c(o, e)$$

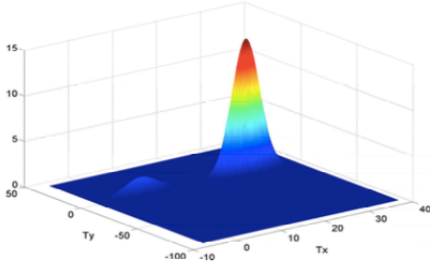
où  $C$  est l'ensemble des cliques relatives au voisinage  $\eta$ . Les potentiels  $V_c$  sont définis localement sur des cliques  $c$ , leurs expressions analytiques dépendent des propriétés locales et globales du problème posé, les  $V_c$  permettent de définir pour chaque nouveau critère la probabilité pour un site donné  $s$  d'être étiqueté par  $e$ . Dans notre cas l'énergie s'exprime sous la forme :

$$U(o, e) = \sum_{i=1}^5 V_{c_i}(o, e) = \sum_{i=1}^5 \alpha_i \cdot W_i(o, e) \quad (7.1)$$

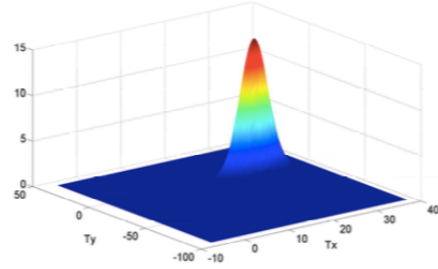
où chaque  $\alpha_i$  est le poids de l'énergie élémentaire  $W_i$  qui représente un des 5 critères retenus.

### Connexité spatiale

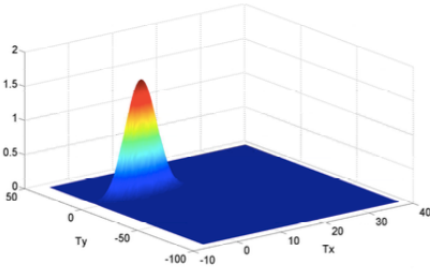




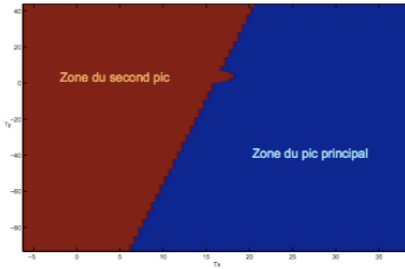
(a) Espace d'accumulation.



(b) Mode principal (mouvement global).



(c) Second mode (mouvement local).



(d) Espace d'accumulation segmenté.

FIGURE 7.9 – Segmentation de l'espace d'accumulation des vecteurs mouvement non compensés (extrait de [217]).

Au sein du segment, nous favorisons les objets (agglomérats de tubes) homogènes et compacts. L'énergie élémentaire, associée à un système de cliques  $\eta_s$  d'ordre 2 (c'est-à-dire un voisinage 8-connexe), est :

$$W_1(e_s) = \sum_{c_s \in C_s} V_{c_s}(e_s, e_t)$$

avec la fonction potentielle :

$$\forall t \in \eta_s \begin{cases} V_{c_s}(e_s, e_t) = \beta_s \text{ si } e_t \neq e_s \\ V_{c_s}(e_s, e_t) = -\beta_s \text{ si } e_t = e_s \end{cases} \quad (\text{où } \beta_s = 1/8)$$

### Homogénéité des couleurs

Au sein du segment, nous attendons un organisation cohérente au niveau de la distribution des couleurs des objets. Pour mesurer la similarité entre couleurs d'un site et celles des différents objets existants, nous utilisons le coefficient de Bhattacharyya. Soit  $\hat{s} = \{\hat{s}_u\}_{u=1\dots m}$ , les densités de probabilité discrètes des  $m$  couleurs du site courant  $s$ , et soit  $\widehat{R}(e_s) = \{\widehat{R}(e_s)_u\}_{u=1\dots m}$  celles (de l'objet ou) de la région  $R(e_s)$  constituée des

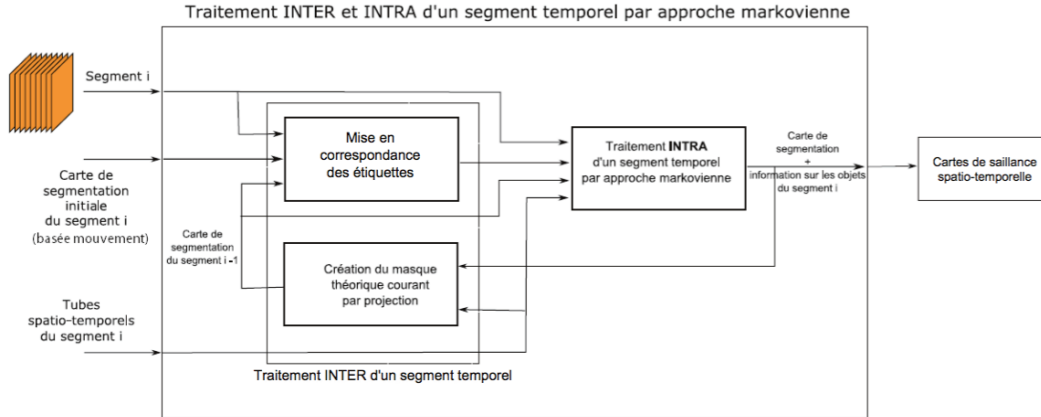


FIGURE 7.10 – Traitement Intra/Inter d’un segment temporel par approche markovienne (extrait de [217]).

sites étiquetés  $e_s$ , le coefficient de Bhattacharyya est défini par :

$$\rho_{\text{couleur}}(\hat{s}, \widehat{R}(e_s)) = \sum_{u=1}^m \sqrt{\hat{s}_u \cdot \widehat{R}(e_s)_u}$$

En pratique, chaque composante couleur est quantifiée sur 16 niveaux (donc  $m = 16^3$ ), et les densités sont calculées à partir d’histogrammes normalisés en considérant les macroblocs de l’image centrale du segment temporel. L’énergie élémentaire « couleur » est :

$$W_2(e_s, o_s, o(R(e_s))) = \sum_{s \in S} V_{\text{couleur}}(\hat{s}, \widehat{R}(e_s))$$

avec la fonction potentielle :  $V_{\text{couleur}}(\hat{s}, \widehat{R}(e_s)) = 1 - 2 \cdot \rho_{\text{couleur}}(\hat{s}, \widehat{R}(e_s))$

### Homogénéité des textures

Au sein du segment, nous attendons aussi un organisation cohérente au niveau des textures des objets. Nous utilisons à nouveau le coefficient de Bhattacharyya pour mesurer la similarité de distribution des textures. En reprenant les mêmes notations,  $\hat{s} = \{\hat{s}_v\}_{v=1..n}$ , sont les densités de probabilité discrètes des  $n$  textures du site courant  $s$ , et  $\widehat{R}(e_s) = \{\widehat{R}(e_s)_v\}_{v=1..n}$  celles (de l’objet ou) de la région  $R(e_s)$  constituée des sites étiquetés  $e_s$ . Les distributions des textures sont évaluées sur l’image centrale du segment, en 2 temps : d’abord en calculant en chaque pixel à l’aide de filtres de Sobel, un couple de gradients (un horizontal et un vertical), puis en quantifiant sur 16 niveaux les gradients (donc  $n = 16^2$ ). Les histogrammes 2D normalisés construits respectivement pour un site et pour une région, sont comparés comme dans le cas de la couleur, et produisent l’énergie élémentaire « texture » :  $W_3(e_s, o_s, o(R(e_s)))$ .

### Homogénéité de mouvement

Au sein du segment temporel, les vecteurs mouvement (chacun associé à un tube) d'un objet donné doivent avoir des valeurs proches. La fonction potentielle mesurant la similarité entre le mouvement d'un site ( $\overrightarrow{VM}_s$ ) et celui d'une région ( $\overrightarrow{VM}_{R(e_s)}$ ) (c'est-à-dire d'un objet) est définie par :

$$V_{\text{mouvement}}(\hat{s}, \widehat{R(e_s)}) = -\frac{\overrightarrow{VM}_s \cdot \overrightarrow{VM}_{R(e_s)}}{\max(\|\overrightarrow{VM}_s\|, \|\overrightarrow{VM}_{R(e_s)}\|)^2}$$

L'énergie élémentaire « mouvement » suit :  $W_4(e_s, o_s, o(R(e_s))) = \sum_{s \in S} V_{\text{mouvement}}(\hat{s}, \widehat{R(e_s)})$

### Connexité temporelle

Il s'agit de suivre les objets pendant la durée de leurs cycles de vie respectifs qui peuvent dépasser un segment temporel. L'opération définie entre 2 segments successifs ( $t-1$ ) et ( $t$ ), peut-être réitérée. Tenant compte des déplacements estimés (mouvement global et mouvements des objets segmentés), la carte de segmentation ( $t-1$ ) est compensée et projetée sur celle du segment courant ( $t$ ). Une étape d'appariement des étiquettes entre objets d'un segment à l'autre, est ensuite réalisée. Elle met en œuvre une métrique basée sur la similarité entre couleurs, celle entre textures, et sur le taux de recouvrement. Pour la couleur ou la texture, un coefficient de Bhattacharyya est adapté. Le cas échéant, un objet du segment courant prend l'étiquette de l'objet dont il est suffisamment proche dans le segment précédent. Les étiquettes des masques d'objets sont alors comparées, la fonction potentielle est définie sur des cliques temporelles :

$$\begin{cases} V_{c_t}(e_s(t-1), e_s(t)) = \beta_t \text{ si } e_s(t-1) \neq e_s(t) & (\text{où } \beta_t = 1) \\ V_{c_t}(e_s(t-1), e_s(t)) = -\beta_t \text{ si } e_s(t-1) = e_s(t) \end{cases}$$

où  $e_s(t)$  (respectivement,  $e_s(t-1)$ ) sont les étiquettes du site du segment courant (respectivement, du segment précédent projeté). L'énergie élémentaire « temporelle » est :  $W_5(e_s(t)) = \sum_{c_t \in C_t} V_{c_t}(e_s(t-1), e_s(t))$  où  $C_t$  est l'ensemble des cliques temporelles de  $S$ .

### Minimisation de l'énergie globale

Il faut minimiser la fonction d'énergie de l'équation 7.1. La carte issue de la segmentation basée mouvement sert à initialiser le processus. Pour chaque site un degré de stabilité est évalué :  $\Delta U(s) = U(s, e_c) - U(s, e_s)$ , il correspond à la différence entre l'énergie associée au site avec l'étiquette courante  $e_c$  et celle qu'il aurait avec une autre étiquette  $e_s$ . Si  $\Delta U(s) > 0$ , alors le site est instable. Nous mettons en œuvre une pile d'instabilité qui procède de façon itérative, en traitant à chaque boucle le site le plus instable (typiquement situé sur le bord de la carte ou d'un objet), jusqu'à ce que tous soient devenus stables. Le traitement d'un site instable est simple : s'il a déjà été examiné, il

est retiré de la liste (car il empêche la solution de converger); sinon il prend l'étiquette qui minimise son énergie et les énergies des sites voisins sont modifiées en fonction. Cette méthode de relaxation déterministe qui converge rapidement vers un minimum local est efficace : elle produit des objets compacts aux bords lissés, et les suit d'un segment à l'autre (voir la figure 7.11)). Une méthode de relaxation stochastique [605, 612, 618, 622] ne serait pas adaptée car trop coûteuse.

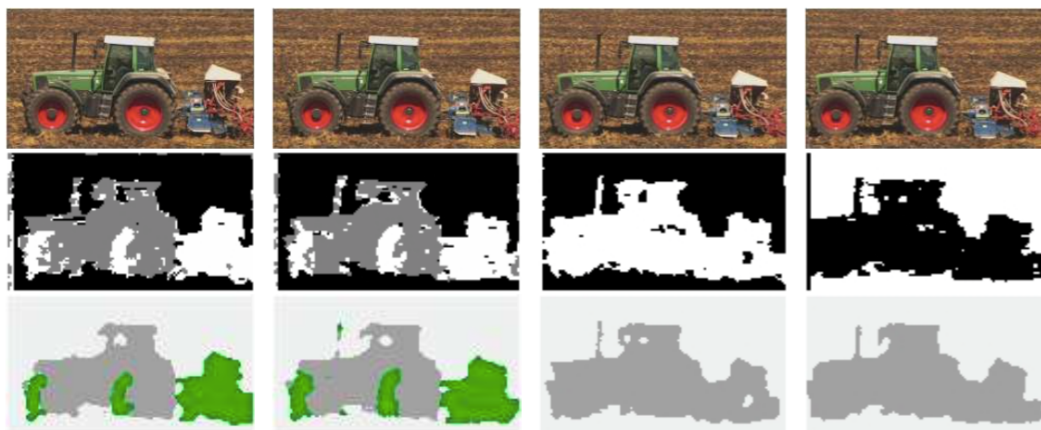


FIGURE 7.11 – Cartes de segmentation pour *Tractor* (segments 13 à 16) : segmentation basée mouvement (ligne du milieu) et approche markovienne (ligne du bas) (extrait de [217]).

### 7.2.3 Carte de saillance visuelle

Nous avons réalisé une modélisation computationnelle de l'AV pré-attentive [38]. Les nombreux facteurs influençant l'AV [436] sont regroupés en deux catégories. La 1<sup>re</sup> regroupe les informations spatiales [267], les plus importantes étant la couleur, l'orientation et la taille. La 2<sup>de</sup> concerne les informations temporelles [611]. Notre modèle d'AV pré-attentive se décompose donc en 2 parties, chacune tenant compte d'un type d'information (spatiale versus temporelle) pour produire une carte de saillance. Une dernière étape combine les 2 cartes afin d'obtenir celle de saillance spatio-temporelle.

#### Saillance spatiale basée sur le contraste de couleur

La littérature [267, 559, 640] souligne l'importance du contraste de couleur contribuant à rendre un objet saillant, on peut distinguer 7 cas :

1. **Contraste de saturation** proportionnel à la différence de saturation. Les couleurs fortement saturées sont généralement saillantes, cas aussi d'une région faiblement saturée entourée par une fortement saturée.

2. **Contraste d'intensité** proportionnel à la différence d'intensité. Les couleurs très lumineuses sont généralement saillantes, cas aussi d'une région sombre entourée par une lumineuse.
3. **Contraste de teinte** proportionnel à la différence des angles de teinte obtenus dans l'espace HSV (voir la figure 7.12), la différence maximale étant de  $180^\circ$ .
4. **Contraste d'opposants** des couleurs en opposition sur le disque de teinte (c'est-à-dire lorsque l'angle est proche de  $180^\circ$ ). Les couleurs de la 1<sup>re</sup> moitié du disque (dite gamme de couleur active) dominent le reste des couleurs dites passives.
5. **Contraste entre couleurs chaudes et froides** dû à la différence entre couleurs chaudes (c'est-à-dire rouge, jaune et orange, situées dans les 1<sup>ers</sup>  $45^\circ$  du disque de teinte) et les autres (dites froides).
6. **Dominance des couleurs chaudes** qui sont plus saillantes.
7. **Dominance de la luminosité et de la saturation** Les couleurs fortement lumineuses et saturées sont plus attractives (quelque soit la valeur de teinte).

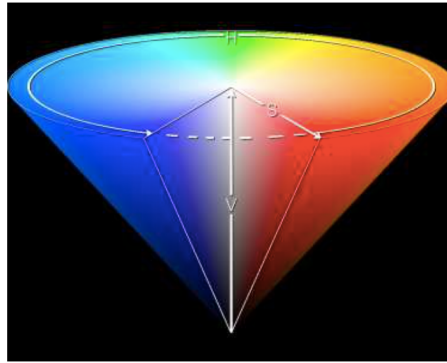


FIGURE 7.12 – Représentation des couleurs dans l'espace conique HSV (ou espace « Teinte, Saturation, Intensité ») (extrait de [38]).

Le calcul de la carte tient compte de ces 7 cas, et suit 7 étapes correspondantes, chacune contribuant à établir la valeur de saillance d'un site  $s$ . Les valeurs des composantes couleur d'un site sont celles moyennes du bloc situé au centre du tube. Les cinq 1<sup>res</sup> étapes utilisent les facteurs de saturation  $f_{ij}^{\text{sat}}$  et d'intensité  $f_{ij}^{\text{int}}$ , où  $i$  et  $j$  représentent respectivement la position du site courant et celle de site voisin :

$$f_{ij}^{\text{sat}} = \frac{S(s_i) + S(s_j)}{2} \cdot (k_{\text{min}} + (1 - k_{\text{min}}) \cdot S(s_i))$$

$$f_{ij}^{\text{int}} = \frac{I(s_i) + I(s_j)}{2} \cdot (k_{\text{min}} + (1 - k_{\text{min}}) \cdot I(s_i))$$

$S(s_i)$  (respectivement,  $I(s_i)$ ) est la valeur de saturation (respectivement, d'intensité) du site  $s_i$ ,  $k_{min}$  est empiriquement introduit pour borner les facteurs, et la valeur maximale de la saturation ou de l'intensité est égale à 1. La 1<sup>re</sup> étape calcule :

$$X_1(s_i) = \sum_{j=1}^{p_i} f_{ij}^{sat} \cdot f_{ij}^{int} \quad \forall s_j \in \eta_i$$

où  $\eta_i$  définit un voisinage 8-connexe à  $s_i$ , et  $p_i$  la taille de ce voisinage. La 2<sup>e</sup> étape calcule :

$$X_2(s_i) = \sum_{j=1}^{p_i} f_{ij}^{sat} \cdot f_{ij}^{int} \cdot \Delta_{ij}^{teinte} \quad \forall s_j \in \eta_i$$

où  $\Delta_{ij}^{teinte}$  représente la différence des teintes des 2 sites :

$$\Delta_{ij}^{teinte} = \begin{cases} \Delta_{ij}^{\mu} & \text{pour } \Delta_{ij}^{\mu} = |H(s_i) - H(s_j)| \leq 0,5 \\ 1 - \Delta_{ij}^{\mu} & \text{sinon} \end{cases}$$

où  $H(s_j)$  est la valeur de teinte (normalisée entre  $[0, 1]$ ) du site  $s_j$ . La 3<sup>e</sup> étape évalue  $X_3(s_i)$  comme  $X_2(s_i)$  mais que si la valeur de teinte en  $s_i$  (respectivement,  $s_j$ ) est active (respectivement, passive), c'est-à-dire si  $H(s_i) < 0,5$  et  $H(s_j) \geq 0,5$ . La 4<sup>e</sup> étape évalue :

$$X_4(s_i) = \sum_{j=1}^{p_i} f_{ij}^{sat} \cdot f_{ij}^{int} \cdot \Delta_{ij}^{sat} \quad \forall s_j \in \eta_i$$

où  $\Delta_{ij}^{sat}$  représente la différence de saturation entre les 2 sites. La 5<sup>e</sup> étape estime :

$$X_5(s_i) = \sum_{j=1}^{p_i} f_{ij}^{sat} \cdot f_{ij}^{int} \cdot \Delta_{ij}^{int} \quad \forall s_j \in \eta_i$$

où  $\Delta_{ij}^{int}$  représente la différence d'intensité entre les 2 sites. Ces 5 premières contributions sont normalisées :

$$V_u(s_i) = \frac{X_u(s_i)}{p_i} \quad u \in \{1 \dots 5\}$$

La 6<sup>e</sup> étape calcule :

$$V_6(s_i) = \begin{cases} S(s_i) \cdot I(s_i) & \text{pour } 0 \leq H(s_i) < 0,125 \quad (\text{intervalle des couleurs chaudes}) \\ 0 & \text{sinon} \end{cases}$$

Enfin la 7<sup>e</sup> étape évalue :  $V_7(s_i) = S(s_i) \cdot I(s_i)$ . La carte de saillance spatiale est obtenue en combinant les 7 contributions :

$$S^{SP}(s_i) = \frac{1}{7} \sum_{u=1}^7 V_u(s_i)$$

Elle est normalisée en fonction de  $S_{max}$  la saillance maximale théorique :  $S^{SP'}(s_i) = \frac{S^{SP}(s_i)}{S_{max}}$

### Saillance temporelle

Le contraste de mouvement est déterminant [611]. A partir de la connaissance du mouvement apparent global  $\vec{V}_\theta$ , et du déplacement apparent local du site (c'est-à-dire son tube)  $\vec{V}_{\text{local}}$ , le vitesse relative du site est évaluée :

$$\vec{V}_{\text{rel}}(s) = \vec{V}_\theta(s) - \vec{V}_{\text{local}}(s)$$

L'indice de saillance temporelle  $S^T(s)$  est obtenu à partir de la fonction illustrée à la figure 7.13 qui tient compte des capacités de poursuite de l'oeil [538].

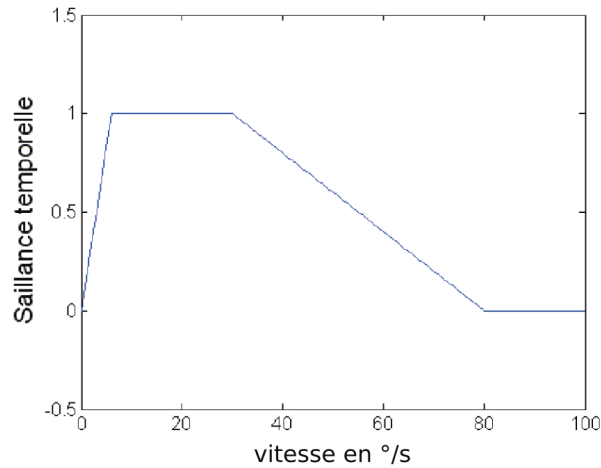


FIGURE 7.13 – Saillance temporelle fonction de la vitesse relative (extrait de [38]).

### Saillance spatio-temporelle

La saillance spatio-temporelle est calculée en combinant saillance temporelle et saillance spatiale :

$$S^{\text{SP-T}}(s) = (S^T(s) + \frac{1}{2}S^{\text{SP}'}(s)).\textit{gauss2D}(s)$$

où *gauss2D* est une fonction gaussienne bi-dimensionnelle introduite pour tenir compte du biais central [388] (c'est-à-dire la tendance des observateurs à favoriser le centre de l'écran). Le contraste de mouvement influençant plus l'AV, le poids de  $S^T$  est deux fois celui de  $S^{\text{SP}'}$ . Des exemples de cartes obtenues sont à la figure 7.14.

#### 7.2.4 Applications au codage vidéo perceptuel

Nous avons testé deux applications possibles de l'outil de pré-analyse avant encodage de la vidéo en flux H.264/AVC, la 1<sup>re</sup> propose une modification adaptative de la structure du GOP en fonction du contenu spatio-temporel du plan vidéo, la 2<sup>nd</sup>e une codage vidéo avec qualité visuelle différenciée guidée par les cartes de saillance.



FIGURE 7.14 – Cartes de saillance pour *Tractor* (segments 13 à 16), avec de haut en bas : les images originales, les cartes de saillance spatiale, celles de saillance temporelle, puis celles de saillance spatio-temporelle (extrait de [38]).

### Modification adaptative de la structure du GOP

Le codeur vidéo a 3 modes de codage pour les images : les images I (codées en mode intra), les images P (images prédites par compensation du mouvement) et les images B (images bi-prédites). La succession des images I, P et B constitue un GOP (« Group of Pictures ») qui débute toujours par une I. La construction du GOP a une influence directe sur le débit du flux compressé car en moyenne, une image I nécessite plus de bits qu'une P qui nécessite plus de bits qu'une B. Les images I sont aussi les images clés (« keyframes ») servant de référence pour la prédiction par compensation de mouvement. De façon classique, un codeur vidéo utilise une taille de GOP fixe, avec typiquement 1 ou 2 images B entre une I et une P, ou entre 2 images P. Afin de s'adapter aux changements spatio-temporels au sein du plan vidéo, la taille du GOP devrait être dynamique [344, 408, 419, 440, 457, 530, 580], il faut donc détecter ces changements nécessitant l'insertion d'une image I, et augmenter le nombre de B des plans comportant peu de mouvement (ou un mouvement panoramique).

**Adaptation du nombre des images B :** Nous avons d'abord vérifié expérimentalement que la configuration optimale du GOP varie bien en fonction de différents types de contenu vidéo, et nous avons également ainsi construit une vérité de terrain. Ces résultats nous ont conduit à caractériser le contenu d'un segment temporel  $ST$  via 2 indices directement évalués à partir de paramètres issus de la pré-analyse : un indice  $IT(ST)$  mesurant l'activité temporelle comme la moyenne des vecteurs mouvement des tubes,



et l'autre  $IS(ST)$  mesurant l'activité spatiale comme la moyenne des EQM obtenues lors des estimations de mouvement relatives aux tubes. Exploitant la vérité de terrain construite expérimentalement, nous avons défini pour un segment  $ST$ , neuf classes d'activités spatio-temporelles possibles (voir la figure 7.15). Ces classes  $C_i$ , où  $i = \{0, \dots, 8\}$ , partitionnent l'espace  $(IS, IT)$ , et  $i$  correspond au nombre d'images B entre 2 images P-P ou I-P. Les bornes des ellipses délimitant les classes ont été déterminées afin que les activités (temporelle et spatiale) entre 2 images P-P ou I-P soient constantes.

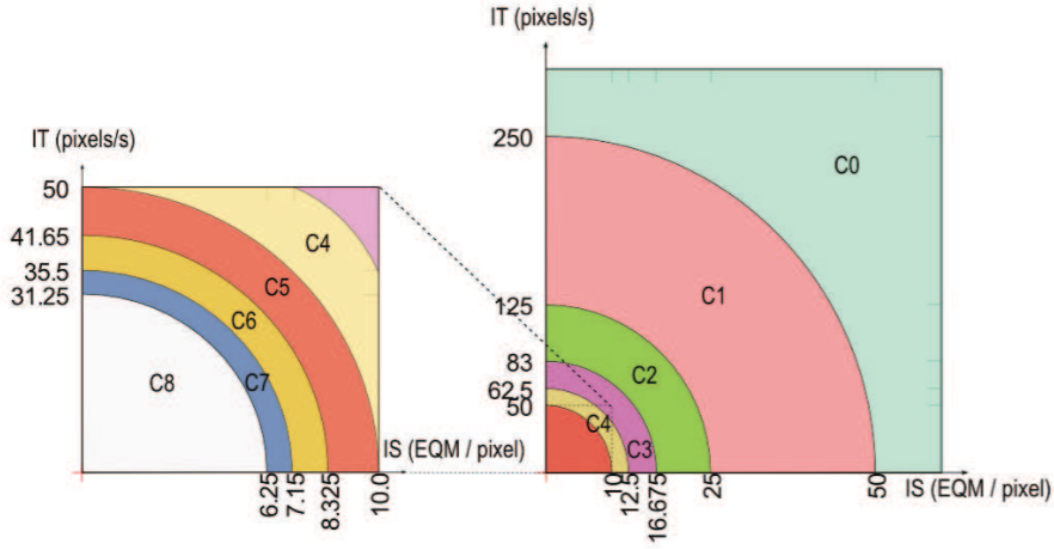


FIGURE 7.15 – Classification des segments temporels en fonction des indices d'activité spatiale  $IS$  et temporelle  $IT$  (extrait de [217]).

Les tests expérimentaux ont comparés 2 codeurs x264 [377] (une version originale, et une version modifiée capable d'adapter le nombre de B), pour 4 débits cibles, et avec (parmi les paramètres de codage) une taille du GOP fixée à 25 images, et 2 images B entre deux images P-P ou I-P. Les résultats moyens sur 8 séquences vidéos, et calculés avec la métrique de Bjontegaard [494] montrent, une réduction du débit de 2,91% et un gain en PSNR de 0,11 dB, les bénéfices sont nets avec les vidéos à fort mouvement.

**Adaptation de la taille du GOP :** Pour savoir quand introduire une nouvelle image I, l'analyse exploite l'indice d'activité temporelle  $IT(ST)$ . Nous avons déterminé empiriquement 2 seuils, un haut  $s_h$  et un bas  $s_b$ , permettant de distinguer 3 cas : Si  $IT(ST) > s_h$ , alors  $ST$  contient un fort mouvement et il faut introduire un nouvelle image I (plusieurs successivement si nécessaire) ; Si  $IT(ST) < s_b$ , alors  $ST$  contient peu de mouvement et il faut espacer davantage les I ; Si  $s_b > IT(ST) < s_h$ , alors  $ST$  contient un mouvement modéré (cas le plus fréquent), et il faut conserver la taille classique du GOP.

Les tests expérimentaux ont comparés 2 codeurs x264 (version originale, versus ver-

sion modifiée capable cette fois d'adapter le nombre de B et l'insertion de I), les autres paramètres de codage sont ceux décrits au paragraphe précédent. Les résultats moyens sur les 8 séquences vidéos, et calculés avec la métrique de Bjontegaard montrent, une réduction du débit de 4,45% et un gain en PSNR de 0,16 dB, les bénéfices sont toujours nets avec les vidéos à fort mouvement.

### Codage adaptatif basé saillance visuelle

Nous développons un schéma de compression sélective directe [573] exploitant les cartes de saillance visuelle  $S^{SP-T}$ . L'objectif est de contrôler la distribution des ressources binaires en fonction de l'intérêt visuel de chaque macrobloc afin d'accroître globalement la qualité visuelle perçue. Comme dans la plupart des cas [443], le paramètre sur lequel on agit est la consigne de quantification : un macrobloc présentant un intérêt visuel faible sera quantifié plus grossièrement qu'un ayant un intérêt visuel important.

**Modification de la carte de saillance :** Il est d'abord nécessaire d'adapter la carte de saillance  $S^{SP-T}$  aux contraintes du codeur H.264/AVC qui permet l'utilisation d'un pas de quantification par macrobloc, seule la variation du pas de quantification entre 2 macroblocs est codée. La simplification et l'homogénéisation de la carte s'effectuent en réalisant sa quantification sur 8 niveaux suivie d'un filtrage morphologique (combinaison d'une ouverture puis d'une fermeture avec un élément structurant  $3 \times 3$ ). La carte de saillance modifiée indique la saillance de chaque macrobloc de l'image centrale du segment temporel. Les cartes de saillance respectives des autres images du segment (c'est-à-dire des 4 précédentes et 4 suivantes) sont déduites à partir des informations de mouvement (projection temporelle dans le sens du mouvement des tubes de la carte de saillance centrale).

**Modification du cœur de codage :** L'indice de saillance calculé pour un macrobloc varie entre 0 (saillance nulle) et 1 (très saillant). Afin de quantifier les macroblocs en fonction de leur indice de saillance, on module la valeur initiale du pas de quantification du codeur (voir la fonction de modulation à la figure 7.16). Si le macrobloc est saillant, on diminue la valeur initiale du pas, dans le cas contraire, on conserve ce pas initial. La consigne globale de débit est contrôlée par l'optimisation débit-distorsion du codeur.

Les tests expérimentaux ont comparés 2 codeurs x264 (version originale, versus version modifiée capable cette fois d'adapter le pas de quantification en fonction des cartes de saillance), les autres paramètres de codage sont ceux décrits aux paragraphes précédents. Les résultats moyens sur les 8 séquences vidéos, et calculés avec la métrique de Bjontegaard montrent : sur les séquences entières, une « réduction » du débit de -0,52% et un « gain » en PSNR de -0,02 dB ; sur les régions d'intérêt de ces séquences, une réduction du débit de 1,75% et un gain en PSNR de 0,06 dB. Les bénéfices sont toujours nets avec les vidéos à fort mouvement.

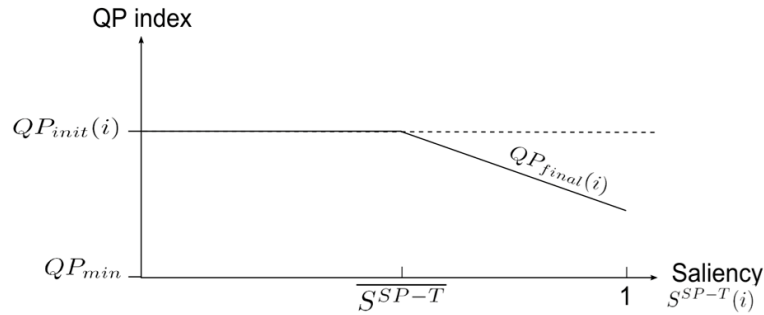


FIGURE 7.16 – Fonction de modulation du pas de quantification du macrobloc en fonction de son indice de saillance,  $S^{SP-T}$  est la valeur moyenne des indices de l'image à laquelle appartient le macrobloc  $i$  (d'après la présentation de soutenance de thèse [217] disponible sur HAL).

Une campagne de tests subjectifs a été menée avec des observateurs « naïfs », afin d'évaluer la qualité visuelle des 8 vidéos codées/décodées via 4 codeurs x264 (version originale, et versions modifiées capables d'adapter la structure du GOP et/ou le pas de quantification en fonction des cartes de saillance). Le protocole ACR (« Absolute category rating ») a été utilisé, afin d'obtenir la note de qualité absolue de chaque séquence (c'est-à-dire le MOS « Mean Opinion Score ») sur une échelle de qualité à cinq catégories [314, 434, 550]. La faible valeur obtenue de  $\Delta Q_{Moy} = Q_{QA} - Q_{x264} = 0,04$ , faisant la différence entre  $Q_{QA}$ , la note de qualité moyenne obtenue avec stratégie de codage, et  $Q_{x264}$  celle obtenue avec le codeur original, montre que nos implantations n'ont pas (encore) apporté d'améliorations significatives.

### 7.3 Discussion

Notre outil de pré-analyse avant encodage vidéo met en œuvre une segmentation spatio-temporelle et un suivi des objets extraits, ainsi qu'un modèle d'attention visuelle pré-attentive. L'unité d'analyse est le segment temporel (dont la durée est liée au temps de fixation du SVH), l'approche innove en se basant sur le tube spatio-temporel pour estimer le mouvement, première des caractéristiques retenues pour extraire l'information pertinente visuellement, et pour conduire un codage cohérent des objets vidéos.

L'outil fournit un ensemble d'informations haut et bas niveaux susceptibles d'être exploitées par de nombreuses applications de codage, nous en avons mis en œuvre deux. La 1<sup>re</sup> avec l'adaptation de la structure du GOP (par variation dynamique du nombre d'images B, et insertion d'images I), en fonction de caractéristiques spatio-temporelles du segment vidéo évaluées via deux indicateurs. Cette application a produit de bons résultats avec des critères d'évaluation classiques (baisse du débit, hausse du PSNR), et en particulier lorsque la vidéo contient de forts mouvements. La 2<sup>nd</sup>e application de compression vidéo avec qualité visuelle différenciée guidée par les cartes de saillance, n'a

pas donné de résultats convaincants (pas d'amélioration de la qualité visuelle des vidéos décodées). La modulation du pas de quantification que nous avons proposée a des effets limités, car la régulation débit-distorsion est contrôlée *in fine* par le codeur. Le verrou demeure toujours l'absence d'une métrique de qualité perceptuelle, susceptible d'être embarquée par le codeur.

Nous verrons dans le chapitre 8 suivant, comment le modèle computationnel de l'attention visuel peut être étendu de la vidéo 2D à la 3D.

## Chapitre 8

# Modélisation computationnelle de l'attention visuelle 2D & 3D

### 8.1 Problématique

Fin 2009, avec Junle Wang et Patrick Le Callet, nous débutons le travail relatif à la « Tâche 1 : Modélisation perceptuelle 2D & 3D » du projet PERSEE (voir le chapitre 4).

Nous avons déjà présenté au chapitre 7 (sections 7.1 et 7.2.3), le mécanisme de l'attention visuelle (AV) déployé par le SVH pour réduire la complexité de l'analyse de la scène visuelle [522]. Des efforts considérables consacrés à l'étude de l'AV, ont mobilisé pendant des décennies différentes disciplines (neurologie, psychologie, vision par ordinateur, etc.). En 2009, c'est la mise en oeuvre du mécanisme de l'AV par des systèmes de traitement d'image qui a un intérêt croissant, car c'est un moyen efficace pour améliorer des applications, par exemple de reciblage [187], de recherche d'images [471], d'évaluation de la qualité [196], ou (comme nous l'avons vu au chapitre 7 précédent) de compression [217, 486].

Nous sommes confrontés à différentes problématiques, et la 1<sup>re</sup> est relative à l'établissement de vérités de terrain 2D. Du fait du lien évident entre AV et mouvements des yeux, les cartes de densité de fixation (CDF) obtenues à partir d'expériences oculométriques, sont considérées comme des vérités de terrain fiables. Cependant, il existe différentes CDF produites via des démarches expérimentales indépendantes. La question se pose au sujet de leur fiabilité, et de leur éventuelle dépendance aux expériences menées. Un 2<sup>e</sup> problème, également relatif à l'établissement de vérités de terrain 2D, est lié à la co-existence en AV de 2 concepts différents aboutissant à désigner des zones visuellement pertinentes. Avec d'un côté la « saillance (visuelle) » issue d'un mécanisme « bottom-up », et de l'autre l'« importance (visuelle) » issue d'un mécanisme « top-down ». Les 2 sont donc conduits par des stimuli de types différents, traités par des voies visuelles et des zones cérébrales différentes. Les 2 concepts sont cependant souvent considérés comme synonymes, nous voulons mieux les caractériser. Le 3<sup>e</sup> problème au-

quel nous nous adressons, est l'impact de la 3D dans le domaine de l'AV. En effet, si un grand nombre de modèles computationnels de l'AV pour les images fixes et la vidéo 2D existe, il n'y en a pas pour la 3D. Nous voulons donc produire un tel modèle. Il faut tenir compte de la perception de la profondeur qui modifie le comportement de la vision humaine [187, 232]. Il faut penser à construire une base de données oculométriques d'images 3D, afin d'évaluer quantitativement les résultats et comparer les performances des différents modèles. Enfin il faut concevoir de nouvelles applications tirant parti de l'AV3D, nous en proposons une adressant le problème de l'inconfort visuel lié à la TV 3D.

## 8.2 Travaux menés et résultats

### 8.2.1 Comparaisons de vérités de terrain de l'attention visuelle 2D

#### Comparaison de CDF 2D issues de bases indépendantes

**Comparaison directe des CDF 2D :** La 1<sup>re</sup> étude porte sur la comparaison directe de 3 bases de CDF 2D issues d'expériences conduites dans 3 laboratoires<sup>1</sup> (à Nantes à l'IRCCyN, à Delft [254], et à Sydney [249] (voir la figure 8.1). Les 3 expériences varient essentiellement sur les observateurs (nombre de 15 à 21, différence culturelle), et les conditions d'observation (distance d'observation, taille de l'écran, durée d'observation entre 10 s et 15 s, point de fixation central entre 2 stimuli pour Sydney). Deux mesures de similarité sont utilisées pour comparer les CDF après seuillage et binarisation : le coefficient de corrélation linéaire de Pearson (CCLP), et l'aire sous la courbe ROC (AROC). Les CDF ont finalement été jugées très semblables, leur degré de similarité dépend néanmoins du contenu de l'image, les images ne contenant qu'une seule région saillante distincte montrent des similarités plus grandes que les autres (c'est-à-dire avec plusieurs régions saillantes ou aucune).

**Comparaison avec des modèles de saillance :** La 2<sup>e</sup> étude porte sur la comparaison des CDF avec des cartes de saillance issues de 5 modèles computationnels bien connus [245, 247, 289, 316, 542]. Les mesures de similarité sont toujours CCLP et AROC. Les résultats varient, cela est dû aux différences initiales existantes entre les cartes des modèles.

**Evaluation de la qualité d'images :** Pour la 3<sup>e</sup> étude, les CDF sont intégrées au sein de métriques de qualité d'image. La CDF pondère alors la carte des distorsions

---

1. Exactement les travaux ont été conduits dans le cadre d'une coopération internationale entre l'Université de Nantes (France), l'Université Technologique de Delft (Pays-Bas), l'Université de l'ouest de Sydney (Australie) et les laboratoires de recherche de Philips (Pays-Bas).

locales estimées, l'opération vise à améliorer les performances de la métrique objective [196]. Trois métriques sont utilisées : PSNR [372], SSIM [435], et VIF [362]. Les résultats montrent que, quelque soit la classe de distorsion, les améliorations apportées par les CDF sont en moyenne très proches, cela traduit leur cohérence (même si les résultats des métriques de qualité le sont moins).

**Reciblage basée sur la saillance :** Une 4<sup>e</sup> étude intègre les CDF dans des algorithmes de reciblage d'image [302], ceux-ci redimensionnent l'image en lui ôtant des bandes verticales de basse « énergie », préservant ainsi la région importante. La CDF automatise le processus en désignant la région saillante importante [205]. Nous avons donc utilisé les différentes CDF, et comparé les similarités entre images après reciblage. Les résultats sont comparables.

### Relation entre mécanismes ascendants et descendants de l'attention visuelle 2D

Nous avons déjà expliqué à la section 7.1 du chapitre 7, comment l'AV repose sur deux processus interagissants, un mécanisme pré-attentif « bottom-up » (ascendant), et un mécanisme attentif « top-down » (descendant). La carte d'importance visuelle [279, 510, 544, 571] associée à une image qui a été préalablement segmentée en objets, indique l'importance visuelle de ces derniers. Les notes d'importance données par des observateurs, sont considérées comme résultantes d'un mécanisme « top-down » car les décisions faites pour évaluer les objets impliquent l'interprétation de la scène, la reconnaissance des objets, et des considérations d'ordre esthétique. Les trajectoires oculaires enregistrées à l'aide d'un oculomètre lors de l'observation libre d'une image, sont considérées (au moins à leur début) comme résultantes d'un mécanisme « bottom-up ». Exactement les fixations désignent les régions saillantes de l'image [542, 613] qui se distinguent par un ou plusieurs attributs (contraste de luminosité, forme, couleur, orientation, etc.), et les cartes de saillance (c'est-à-dire les CDF) exploitent leur densité. Les 2 types de cartes (importance versus saillance) permettent donc de désigner les parties de la scène visuellement « pertinentes ». Elles sont souvent considérées comme synonymes malgré les différences qui les déterminent en termes de traitement visuel. Nous voulons donc établir les relations entre les deux.

**Expériences psychophysiques conduites :** A des fins de comparaison, deux expériences psychophysiques ont été menées afin de construire les 2 types de cartes pour 80 images de la même base (celle de Berkeley contenant images et cartes de segmentation).

**Analyse qualitative des cartes :** Qualitativement les cartes présentent similitudes et différences (voir des exemples à la figure 8.2). Les cartes importance montrent que la catégorie de l'objet (par exemple les visages ou les animaux) joue un rôle déterminant dans la notation. Les cartes de saillance montrent que les régions possédant

une caractéristique spatiale locale remarquable attirent l'attention, et que les observateurs regardent toujours le sujet principal (et moins l'arrière-plan). Les CDF ne séparent évidemment pas en objets.

**Prédiction du sujet principal et des secondaires :** Nous voulons prédire les cartes d'importance à partir des cartes de saillance (en utilisant aussi les cartes de segmentation). Deux approches ont été testées pour calculer la valeur d'importance d'un objet : l'une calcule sa « saillance moyenne » (pour l'objet, les valeurs de saillance des pixels sont sommées, et le total divisé par le nombre de pixels), l'autre sa « saillance de couverture » (cette fois le total n'est divisé que par le nombre de pixels ayant une saillance supérieure à 0). Les objets sont classifiés en 3 classes d'importance en fonction de leur valeur d'importance : un « sujet principal » a une valeur dans l'intervalle  $[\frac{2}{3}; 1]$ , un « objet secondaire » a la sienne dans  $[\frac{1}{3}; \frac{2}{3}]$ , et un « arrière plan » dans  $[0; \frac{1}{3}[$ . L'analyse des « matrices de confusion » donnant les taux de succès (chaque ligne représente la classe d'importance réelle des objets et chaque colonne la classe prédite), montre que la saillance moyenne prédit avec succès le sujet principal dans 81% des cas, et que la saillance de couverture, moins performante à prédire les sujets principaux, est meilleure (67% de succès) pour les arrière-plans.

**Analyse temporelle :** Nous comptons par intervalles de 100 ms, en considérant toutes les observations de 15 s des images, le nombre de fixations oculaires par classe d'importance (voir la figure 8.3). Ces trois courbes montrent que la relation entre saillance et importance visuelle est dépendante du temps. Pendant l'intervalle  $[0; 0,5]$  s, les mécanismes « bottom-up » sont effectifs, et le nombre de fixations sur les sujets principaux est déjà 4 à 6 fois plus élevé que pour les autres types d'objets. Entre  $[0,5; 2]$  s, les fixations demeurent fortes sur les sujets principaux que les observateurs analysent, c'est l'influence de mécanismes « top-down ». Ensuite le nombre de fixations vers les objets des 2 autres classes augmente, car les observateurs explorent la totalité de l'image (même si leur attention va toujours vers les sujets principaux). Le fait que les sujets principaux attirent plus l'attention pendant les 2 premières secondes d'observation indique que ces premières fixations peuvent mieux prédire cette classe. Nous avons donc recalculé la « matrice de confusion » en n'utilisant que ces premières fixations, et la saillance moyenne a prédit avec succès le sujet principal dans 89% des cas,

## 8.2.2 Modélisation computationnelle de l'attention visuelle 3D

### Prise en compte du biais de profondeur

Lors de la visualisation d'images ou de vidéos 2D, le « biais central » (ou « biais de fixation au centre ») a été mis en évidence [263, 333]. Lors de la visualisation de contenus 3D (sur écran plat stéréoscopique), il a été montré que les fixations des observateurs sont biaisées, non seulement vers la zone centrale de l'écran, mais aussi vers certains plans de profondeur de la scène [18, 252, 260]. Pour l'élaboration de modèles computationnels



de l'AV en visualisation stéréoscopique, des hypothèses de « biais de profondeur » ont été proposées [221, 243, 514, 560], cependant les études psychophysiques de ce biais font encore défaut.

**Expérience psychophysique conduite :** Nous avons donc mené une expérience oculométrique binoculaire conduite en montrant des stimuli synthétiques sur un écran stéréoscopique récent. Les scènes présentent des objets identiques affichés à différents plans de profondeur (voir la figure 8.4), la disparité est le seul indice de profondeur utilisé. Les stimuli ont aussi été conçus de façon à éviter le plus grand nombre possible de caractéristiques visuelles 2D. Vingt-sept observateurs ont participé à l'expérience. Après filtrage des saccades, les fixations sur les objets ont été identifiées. Les profondeurs de ces derniers étant connues, les profondeurs respectives des fixations en ont été déduites.

**Résultats :** Les résultats expérimentaux montrent clairement que les observateurs prêtent plus, et plus tôt, attention aux objets les plus proches d'eux. Ce phénomène pourrait être dû à une stratégie visuelle ancestrale selon laquelle l'homme explore la scène à partir des objets les moins distants qui peuvent représenter dans la nature un danger potentiel [608]. Nous avons trouvé que le biais de profondeur est évident dès le début de l'observation (c'est-à-dire dès la 1<sup>re</sup> fixation de l'observation), le court temps de réaction implique qu'il doit être le résultat d'un mécanisme « bottom-up ».

### Conception d'un modèle de l'attention visuelle 3D basé saillance de profondeur

Un nouveau modèle de l'attention visuelle 3D, s'appuyant à la fois sur les caractéristiques visuelles 2D et les caractéristiques extraites de l'information de profondeur, est proposé.

**Génération de la carte de saillance de profondeur :** La carte de saillance de profondeur est calculée dans un cadre bayésien. Soit  $S$  la saillance de profondeur en chaque pixel de la carte, elle est exprimée comme une probabilité :

$$S = P(C = 1 | \bar{f}_{dep}, l_z)$$

où  $C$  est la variable binaire aléatoire indiquant si le point est regardé ou non, le vecteur aléatoire  $\bar{f}_{dep}$  indique les caractéristiques de profondeur au point, et  $l_z$  sa profondeur dans la scène 3D. En faisant l'hypothèse que  $\bar{f}_{dep}$  est indépendant de la distance à l'observateur, et que  $P(C = 1)$  est une simple constante, alors  $S$  peut s'écrire :

$$S = P(C = 1 | \bar{f}_{dep}) \cdot P(C = 1 | l_z) \cdot \text{const}_1 \quad (8.1)$$

Le 1<sup>er</sup> terme de l'équation 8.1 peut s'écrire d'après Bayes :

$$P(C = 1 | \bar{f}_{dep}) = \frac{P(\bar{f}_{dep} | C = 1)}{P(\bar{f}_{dep})} \cdot \text{const}_2$$

Donc la probabilité du point d'être regardé est liée à la distribution des caractéristiques autour de ce point (normalisé par  $P(\bar{f}_{dep})$ , lié à la « rareté » des caractéristiques de profondeur). Ce 1<sup>er</sup> terme est calculé en utilisant deux informations : le contraste de profondeur (considéré ici comme synthétisant les caractéristiques de profondeur au point), extrait via un filtre de différence de gaussiennes (DOG) appliqué en chaque point de la carte de profondeur ; et la probabilité que le point soit regardé modélisée en se basant sur des données expérimentales (c'est-à-dire les mouvements oculaires recueillies lors de l'expérience menée pour étudier le biais de profondeur, voir le paragraphe précédent). Le 2<sup>e</sup> terme de l'équation 8.1 correspond au biais de profondeur.

**Schéma du modèle d'attention visuelle 3D :** Nous proposons un nouveau modèle computationnel de l'AV pour image stéréoscopique 3D fixe. Le schéma (voir la figure 8.5) combine la carte de saillance de profondeur à celle calculée sur la base de caractéristiques visuelles 2D, pour prédire la carte de saillance 3D finale. Pour le calcul de l'AV 2D, nous utilisons trois modèles existants : le modèle de Itti [542], celui de Bruce [247] et celui de Hou [316]. Nous adoptons une approche directe [199] pour combiner  $SM_{dep}$  la carte de saillance de profondeur, à  $SM_{2D}$  celle de saillance 2D, et produire  $SM_S$  celle finale :

$$SM_S = \omega_1 \cdot SM_{dep} + \omega_2 \cdot SM_{2D} \quad \text{avec } \omega_1 = \omega_2 = 0,5 \quad (8.2)$$

**Création d'une base de données oculométriques de contenu 3D :** Face au manque de vérités de terrain limitant l'étude des modèles de l'AV 3D, nous avons créé une base de données contenant : 18 images naturelles stéréoscopiques (10 images de la base Middlebury [331], et 8 de la base IVC), leurs cartes de disparité, et les données oculométriques capturées lors des tâches des visualisations libres de 35 observateurs « naïfs ». Les points (des yeux gauche et droit) enregistrés par l'oculomètre, ont été fusionnés [187], puis les cartes résultantes filtrées par un noyau gaussien bidimensionnel (tenant compte lorsque augmente l'excentricité depuis la fovéa, des diminutions de l'acuité visuelle, et de la précision de l'oculomètre) afin de produire les cartes de densité de fixation (CDF) finales.

**Evaluation des performances du modèle d'attention visuelle 3D :** La figure 8.6 donne des exemples de cartes de saillance obtenues : cartes de saillance de profondeur prédites par notre modèle, et cartes de saillance 2D prédites par les 3 modèles. La comparaison qualitative se fait avec les CDF (générées à partir des données oculométriques) qui sont la vérité de terrain : nos cartes intègrent mieux les zones saillantes des images 3D. Trois mesures de similarité sont utilisées pour comparer quantitativement les cartes de saillance aux CDF : le coefficient de corrélation (CCLP) [228, 351], la divergence de Kullback-Leibler [247, 351], et l'aire sous la courbe ROC (AROC) [212, 296]. La meilleure performance, de nos cartes de saillance de profondeur sur les cartes de saillance 2D, est objectivement confirmée. Enfin 2 approches, pour combiner la carte

de saillance 2D et la carte de saillance de profondeur, sont quantitativement comparées : la 1<sup>re</sup> procède par multiplication [221], la 2<sup>nd</sup>e est l'approche directe de l'équation 8.2. Les cartes combinant les saillances sont supérieures à celles de saillance 2D seule, les performances de la méthode de combinaison directe sont les meilleures.

### 8.2.3 Amélioration de la qualité d'expérience de visualisation de la TV3D

L'objectif est de proposer une application potentielle du modèle de l'attention visuelle 3D, et nous retenons un contexte de visualisation de contenus 3D stéréoscopiques. Nous voulons montrer qu'il est possible de tirer avantage de nos connaissances (détermination de la région d'intérêt, effets des indices de profondeur sur la perception de la profondeur de la scène 3D), pour améliorer la qualité de l'expérience de visualisation 3D.

#### Perception de la profondeur et conflit en vision stéréoscopique

Le SVH utilise des indices pour estimer la profondeur d'une scène 3D. Ces indices de profondeur sont classés en fonction du nombre d'yeux impliqués [258, 572], et l'on distingue les indices monoculaires (taille de l'image rétinienne, hauteur dans le champ visuel, gradient de la texture, flou de défocalisation, perspective atmosphérique, accommodation, parallaxe de mouvement, ombres et dégradés, occlusion et interposition), de ceux binoculaires (convergence, disparité binoculaire). Il faut remarquer que les indices monoculaires sont aussi disponibles lorsque les 2 yeux sont utilisés.

Le conflit entre les 2 indices de profondeur, accommodation et convergence des yeux, est généralement considéré comme la principale raison de l'inconfort et de la fatigue visuelle lorsqu'il est regardé un contenu sur un écran 3D stéréoscopique [276] (voir aussi la figure 8.7)

#### Amélioration de la profondeur perçue par ajout de flou de défocalisation

Une voie possible pour gagner de la profondeur apparente tout en évitant l'inconfort visuel dû à une disparité binoculaire trop forte, est de renforcer les indices monoculaires de profondeur. Parmi les indices monoculaires de la profondeur, le flou de défocalisation est connu pour être important [234, 235, 409, 614], cependant d'autres études [287, 563, 564] annoncent que le flou n'a, soit aucun effet, soit seulement certains effets qualitatifs sur l'ordonnancement de la profondeur perçue. Nous avons donc conduit une étude montrant comment la combinaison de flou et de la disparité affecte la profondeur perçue.

**Expérimentation :** Chaque stimulus utilisé pour ce test se compose de la superposition d'un seul objet au premier plan (l'image nette d'un papillon) sur un arrière-plan (l'image d'un parterre de fleurs), le cas échéant le fond a été spatialement flouté par un filtre gaussien (voir aussi la figure 8.8). La figure 8.9 illustre le schéma de principe utilisé

pour la construction des stimuli dont seuls 2 paramètres varient : la position absolue ( $D_a$ ) du plan du fond, et la distance relative ( $D_r$ ) de l'objet au 1<sup>er</sup> plan. A un « BB-stimulus » (au fond flouté) caractérisé par un couple ( $D_a, D_r$ ) donné, sont associés 7 ou 8 « SB-stimuli » (au fond net) pour lesquels varie  $D_r$  dans une plage donnée. Les stimuli, sont visualisés binoculairement à travers des lunettes à obturation active. L'expérience subjective est une tâche de choix forcé où les observateurs sont invités à déterminer entre les paires de stimuli affichées, si le « BB-stimulus » présente un intervalle de profondeur plus large (entre le papillon et le fond) que le cas du « SB-stimulus ».

**Analyse des résultats :** Nous mesurons la proportion de réponses « Le BB-stimulus contient un intervalle de profondeur plus large ». Le résultat est tracé comme une fonction de la différence de disparité entre les stimuli « BB » et « SB ». La fonction cumulative de Weibull est utilisée comme fonction psychométrique. La différence de disparité correspondant au point à 50% est considérée comme le point d'égalité subjective (PSE), et la mesure en ce point donne l'augmentation de la profondeur perçue. En traçant la courbe du PSE en fonction de l'intervalle de profondeur entre l'objet au premier plan et l'arrière-plan, nous avons déterminé que la disparité peut effectivement être compensée par le flou, et que l'augmentation de la profondeur perçue est approximativement égale à 300 secondes d'arc (et cela quel que soit le plan de profondeur auquel on ajoute le flou). En outre, nous avons également constaté que le flou compense mieux la disparité s'il est ajouté à un plan de profondeur plus éloigné de l'avant-plan.

Le résultat expérimental montre donc la possibilité d'améliorer par flou de défocalisation la profondeur perçue dans une scène 3D stéréoscopique. C'est là un procédé intéressant permettant d'atténuer le conflit possible entre accommodation et vergence, car nos résultats montrent qu'il est possible de diminuer la disparité binoculaire des objets sans perdre l'effet « pop-out » en ajoutant un effet de flou sur l'arrière-plan.

### 8.3 Discussion

Ces travaux ont portés sur plusieurs aspects de la recherche sur l'attention visuelle, en abordant des problématiques liées aux vérités terrain de l'AV, et celles relatives à la modélisation de l'AV dans des conditions de visualisation 3D. Les résultats ont été nombreux et importants (avec par exemple, notre contribution au tout 1<sup>er</sup> modèle computationnel de l'AV 3D basé sur la saillance de profondeur). Une part considérable du travail a été menée par le biais d'expériences psychophysiques (ce fut d'ailleurs ma 1<sup>er</sup> expérience poussée dans ce type de recherche), et notre but n'était pas ici de développer les applications au codage vidéo (ceci est conforme à l'organisation du projet PERSEE décrite à la partie 4.2 du chapitre 4).

Ces travaux de thèse illustrent de façon typique les activités importantes à conduire, de façon à irriguer la recherche en codage vidéo perceptuel, et particulièrement dans un contexte d'émergence de nouveaux formats vidéo immersifs (UltraHD « Ultra Haute

Définition », HFR « High Frame Rate », HDR « High Dynamic Range », etc.), lorsque les besoins se renforcent. Au delà de la mise au point de modèles perceptuels susceptibles de piloter des approches de compression, il faut établir des vérités de terrain afin de tester les codeurs et comparer leurs résultats. Les bases de données vidéo doivent être riches en formats et en contenus. Ces derniers doivent être organisés selon des taxonomies établies (un outil de codage spécialisé peut être testé sur un type ciblé de contenu), enrichis de métadonnées (index d'activité spatio-temporelle, conditions de capture, descriptions sémantiques, etc.), et les bases doivent demeurer ouvertes et publiquement disponibles. Un autre axe de travail est l'exploitation des modèles perceptuels à des fins d'amélioration de la qualité d'expérience (QoE) de l'utilisateur final du service vidéo (corollairement, on peut ainsi élargir le champ des possibilités des créateurs de vidéos). Le chapitre le montre avec l'exploitation possible du flou de défocalisation pour accroître la profondeur perçue d'une scène 3D stéréoscopique, par ailleurs d'autres propriétés de flous en vidéo (par ex. le flou de mouvement) mériteraient d'être exploitées en codage vidéo perceptuel.

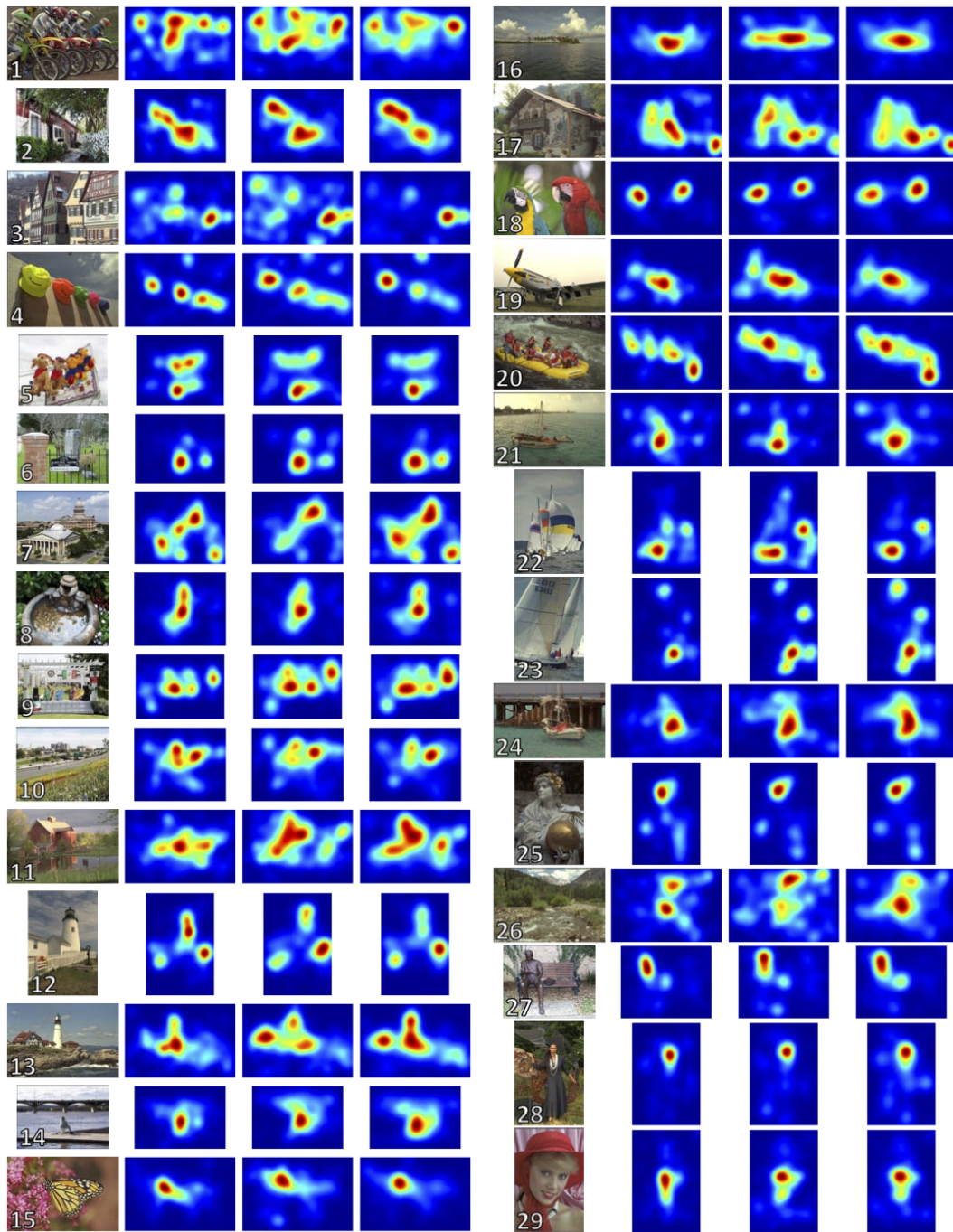


FIGURE 8.1 – Exemples de CDF utilisées, avec de gauche à droite : l'image originale (issue de la base LIVE [402]), puis les 3 CDF (de Sydney, Delft, puis Nantes) (extrait de [171]).

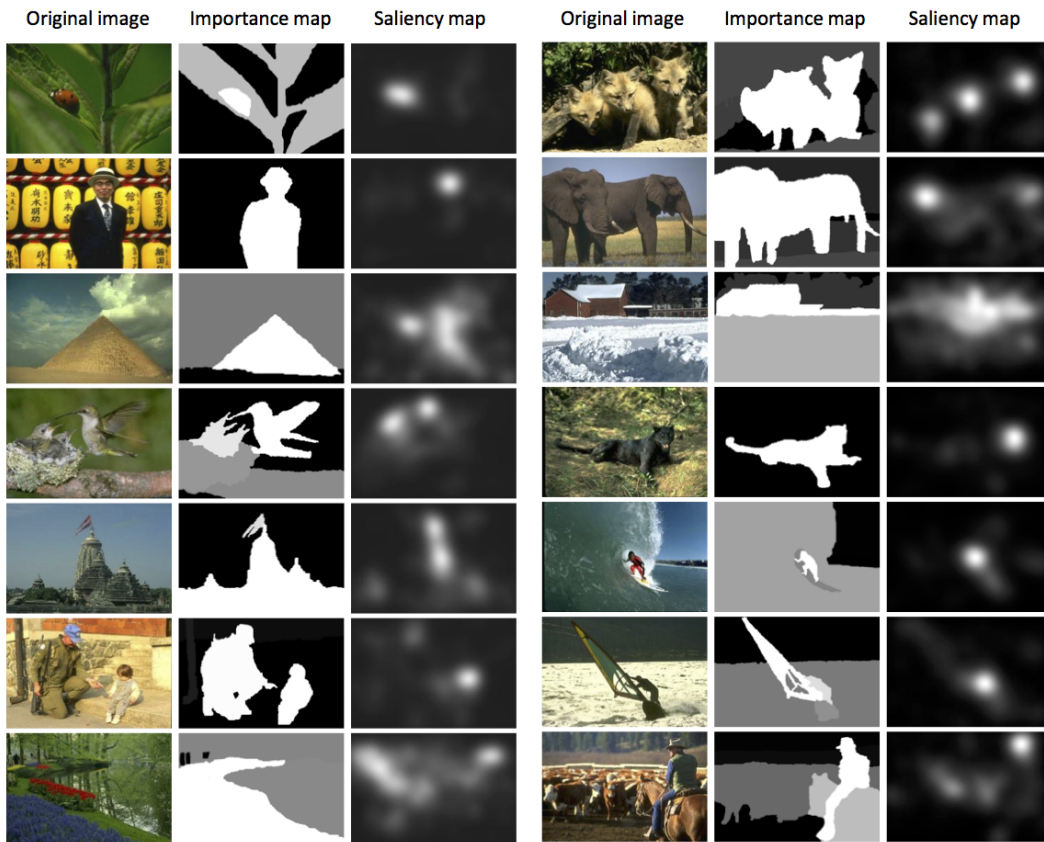


FIGURE 8.2 – Exemples de cartes issues des expériences, avec de gauche à droite : l'image originale, la carte d'importance et celle de saillance (extrait de [242]).

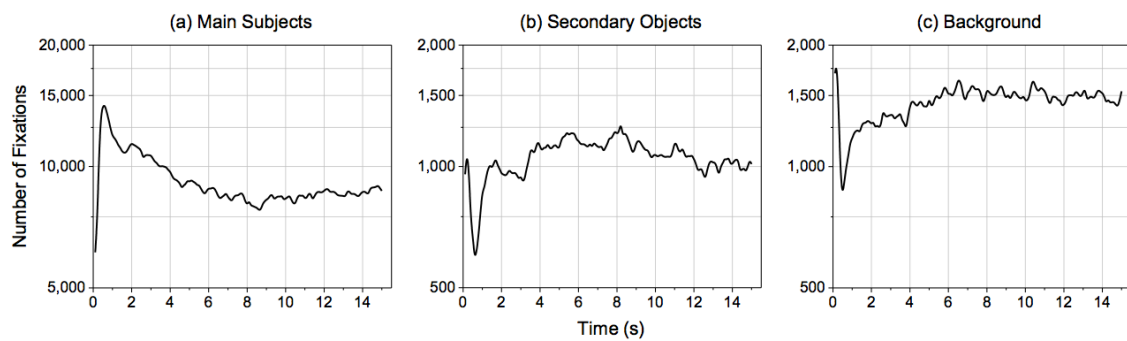


FIGURE 8.3 – Nombre de fixations oculaires sur (a) les sujets principaux, (b) les objets secondaires, et (c) les objets d'arrière-plan, pendant la durée d'observation (extrait de [242]).

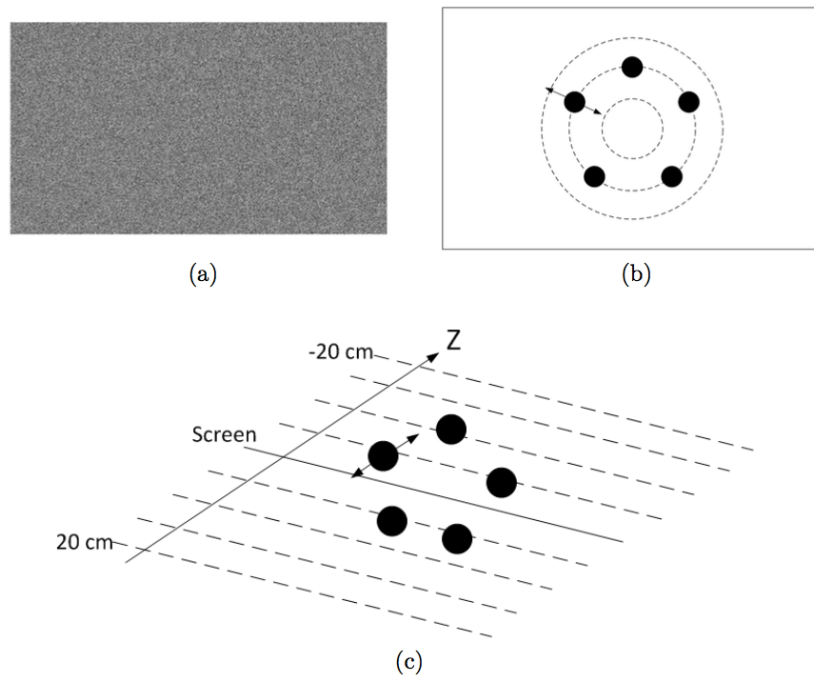


FIGURE 8.4 – Composition des stimuli : (a) arrière-plan (un bruit blanc), (b) positions des projections des objets sur l'écran, (c) distribution des objets en profondeur (extrait de [3]).



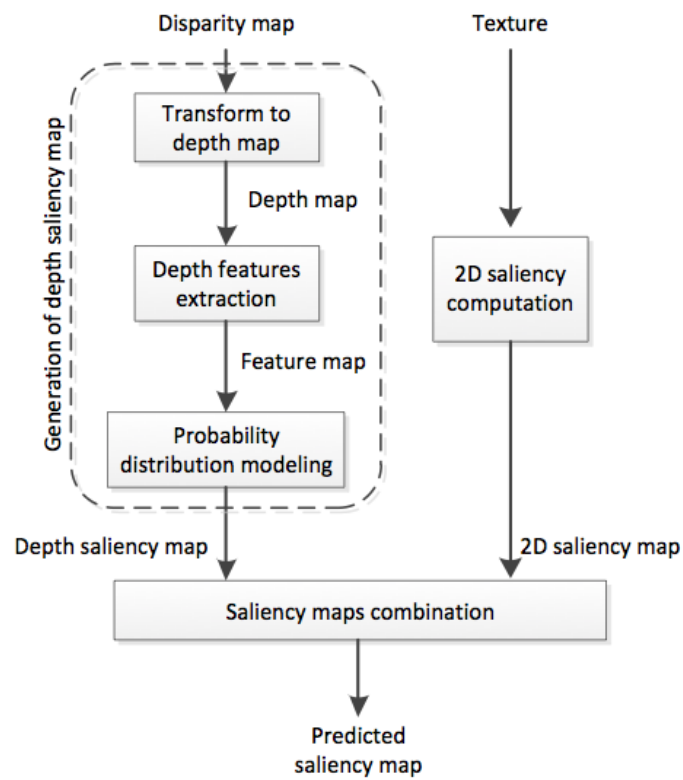


FIGURE 8.5 – Schéma d'ensemble du modèle d'attention visuelle 3D proposé (extrait de [2]).

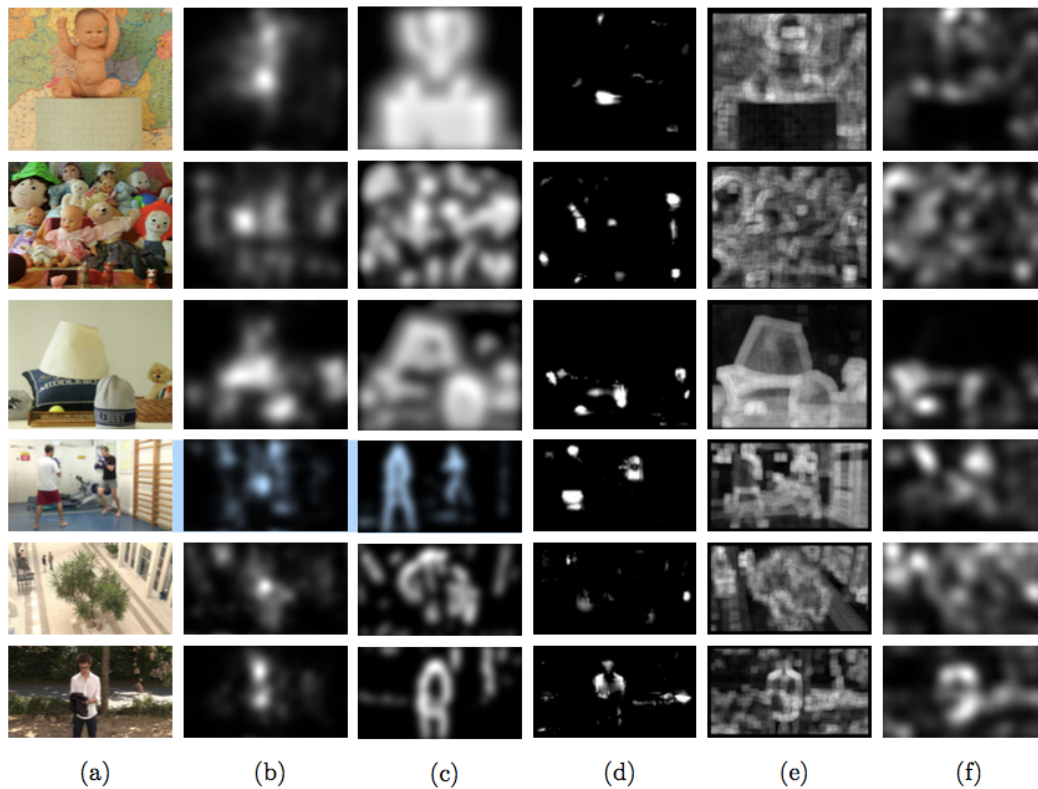


FIGURE 8.6 – Exemples de cartes de saillance obtenues, avec de gauche à droite : (a) l'image originale, (b) la CDF (vérité de terrain), puis les cartes de saillance calculées selon (c) le modèle proposé (saillance de profondeur uniquement), (d) Itti, (e) Bruce, et (f) Hou (extrait de [171]).

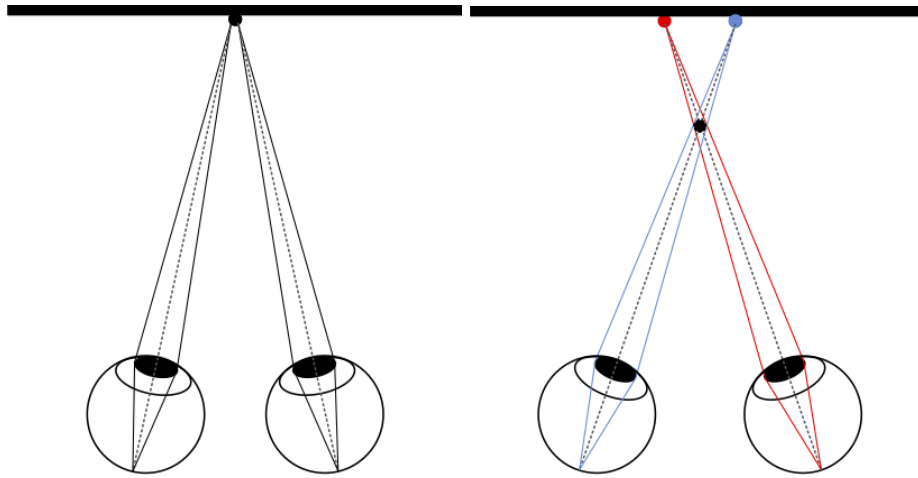


FIGURE 8.7 – Illustration du conflit entre accommodation et vergence. A gauche le cas de visualisation d'un contenu 2D : les yeux s'accommodent et convergent naturellement vers le même objet sur l'écran. A droite le cas de visualisation d'un contenu 3D stéréoscopique : les yeux s'accommodent sur l'écran mais convergent vers un objet ici en avant plan, les distances focale et de convergence différent, en dehors d'une plage de tolérance, inconfort visuel et fatigue visuelle apparaissent (extrait de [171]).



FIGURE 8.8 – Exemples de stimuli : avec à gauche le cas d'un « SB-stimulus » dont le fond est net, et à droite le cas d'un « BB-stimulus » dont le fond a été flouté (extrait de [17]).

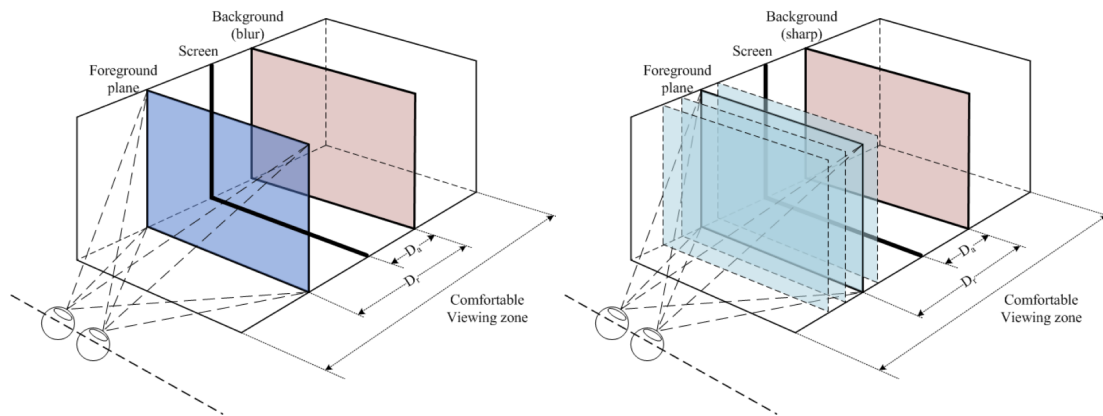


FIGURE 8.9 – Construction des stimuli : les plans de profondeur bleu contiennent le papillon en avant du plan rouge qui est le fond. A un « BB-stimulus » (à gauche), sont associés sept ou huit « SB-stimuli » (à droite) (extrait de [17]).

## Chapitre 9

# Modélisation de la similarité des textures visuelles, et compression

### 9.1 Problématique

Début 2014, avec Patrick Le Callet, nous entamons un travail contribuant au « WP4 Outils perceptuels et évaluation de la qualité » du projet ITN PROVISION. Comme expliqué dans le chapitre 5, le projet a retenu une approche de codage vidéo basé contenu (voir sa description à la figure 5.1) mettant en œuvre une analyse/synthèse des textures [211, 246, 262]. Les objectifs de la thèse ont alors été définis afin de contribuer directement à ce modèle de codage vidéo basé contenu, et propose successivement : d'étudier les modèles de perception des textures spatio-temporelles, d'identifier les représentations possibles compatibles avec le codeur HEVC [165], et de définir une métrique perceptuelle efficace dans ce contexte.

### 9.2 Travaux menés et résultats

#### 9.2.1 Définition et modélisation des textures visuelles

Au début de la thèse, les approches sur la perception des textures statiques et dynamiques ont été passées en revue [102, 121, 422], en les liant aux modèles existants relatifs à leur similarité visuelle. Nous avons constaté qu'il manquait une définition universelle des textures, dans la mesure où celles statiques [265, 326, 387, 413, 456, 485, 517, 549, 607, 635] tendent à être définies différemment comparées aux textures dynamiques (que l'on distingue dans la littérature, comme temporelles [341, 401, 595], dynamiques [103, 226, 248, 380, 453, 502], spatio-temporelles [141], ou en mouvement [64, 114, 119, 470, 472]). Par conséquent, nous avons jugé utile de produire une définition plus juste de ce qu'est une texture visuelle statique ou dynamique :

*Une texture visuelle est un phénomène englobant à la fois les textures statiques et dynamiques. Les textures statiques sont relatives à des régions homogènes de la scène, elles sont généralement composées de petits éléments (appelés aussi « textons » [627]) disposés dans un certain ordre, et qui ne peuvent présenter que des mouvements simples (translation, rotation et zoom). Quant aux textures dynamiques, elles évoluent au cours du temps de façon complexe en combinant mouvements et déformations, mais elles présentent un certain degré de stationnarité dans l'espace et le temps.*

Nous avons remarqué que de nombreuses études approfondies existent sur la perception des textures statiques. Certaines s'appuient sur un modèle simple de discrimination pré-attentive entre textures [620, 627, 634, 639, 645] où le modèle considère les distributions statistiques de caractéristiques locales (couleur, taille, etc.) des textures. D'autres approches reposent sur des modèles plus complexes du cortex visuel qui est sélectif en fréquence et en orientation. Ces modèles de discrimination entre textures mettent donc en œuvre une étape de filtrage multicanal [102, 121, 422, 427, 561, 616, 623, 641].

Cependant nous nous sommes aperçus que la perception des textures dynamiques n'avait été encore que très peu explorée. Nous nous sommes donc orientés vers l'étude de la perception du mouvement par le SVH [198, 545, 556, 617], et un modèle complet est présenté à la figure 9.1. Il montre que c'est principalement le cortex qui est impliqué, avec d'abord la simple détection du mouvement effectuée au niveau de l'aire primaire V1 par des neurones sélectif en fréquence spatiale et temporelle, orientation, direction et vitesse [198, 203, 418, 545]. Au delà de l'aire V1, deux voies sont différenciées et sont relatives aux mécanismes visuels d'ordre supérieur Il y a la voie dorsale qui est responsable de l'analyse du mouvement [75, 88, 92, 198, 269, 358, 430, 545], on la surnomme aussi « voie du où ? ». Il y a ensuite la voie ventrale qui effectue principalement l'analyse des formes, on la surnomme aussi « voie du quoi ? ». Etant donné que par hypothèse les textures perçues n'ont pas de forme c'est la voie dorsale qui nous intéresse principalement, en écartant les mécanismes au delà de la simple perception des textures (par exemple ceux liés à la poursuite [500, 527]).

### 9.2.2 Similarité visuelle entre textures, et compression

Nous avons lié l'étude précédente à la notion de similarité visuelle entre textures. Exactement la revue des modèles de similarité a révélé qu'il en existe globalement quatre catégories, avec des degrés différents d'association à la notion de perception. Ces catégories de modèles de similarité sont basées : transformation, auto-régression, textons ou mouvement.

La modélisation basée transformation en sous-bandes présente un lien direct entre perception et similarité de textures. Ceci est du au fait qu'il existe une transformation similaire dans la zone de traitement cortical V1. L'approche a été retenue aussi bien avec des textures statiques [97, 134, 266, 298, 326, 480, 508, 517, 561, 596], qu'avec des textures dynamiques [115, 141, 148, 226, 248, 489]. Des modèles auto-régressifs (AR et ARMA) ont été largement utilisés, car les paramètres extraits permettent de reconnaître,

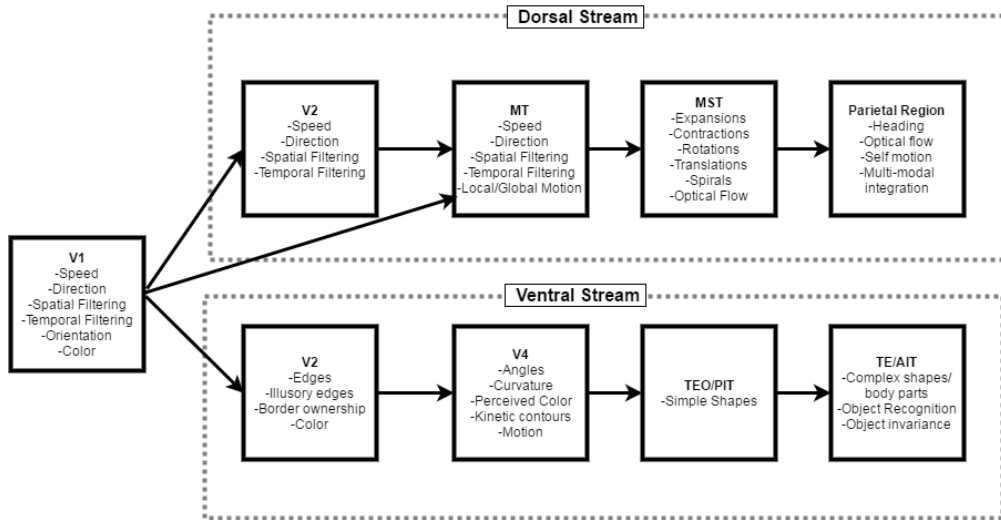


FIGURE 9.1 – Modèle hiérarchique du système visuel, d’après [101].

classifier, segmenter ou synthétiser des textures, aussi bien statiques [107, 281, 414], que dynamiques [94, 272, 376, 439, 453, 454, 482, 502]. Les modélisations basées textons pour les textures statiques [105, 133, 224, 233, 306, 324, 483, 485, 604] et celles dynamiques [336] supposent que la similarité entre textures est directement due à la distribution de leurs textons aux caractéristiques locales remarquables, ces méthodes ignorent donc le traitement neuronal visuel. Enfin des modèles reposent aussi directement sur l’information du mouvement (par exemple après calcul du flot optique [250, 301, 397, 432, 595]) pour caractériser les textures dynamiques, mais là encore il n’y pas de lien effectué avec la perception.

L’exploitation en compression d’images et de vidéos, du concept de similarité entre textures, est le plus souvent effectuée de façon indirecte lorsqu’il n’est pas mis en œuvre au sein de la boucle de codage basé bloc hybride pour améliorer des performances débit-similarité, mais plutôt afin de générer au décodage des textures similaires par synthèse de texture. Trois catégories de codage basé synthèse ont été identifiées, elles reposent sur une suppression, une simplification ou une prédiction des textures.

Les approches basées suppression de texture effacent une grande partie des textures et s’appuient sur la synthèse pour régénérer coté décodeur les parties manquantes [76, 82, 83, 211, 246, 259, 392, 464, 465]. Ces approches sont couramment utilisées car elles peuvent produire une économie en débit significative. Ceci est dû au fait que pour les parties effacées, seuls les paramètres nécessaires à leur synthèse sont encodés, et ils sont beaucoup moins coûteux que les données d’origine. Cependant, deux obstacles empêchent le déploiement de ces modèles : les flux binaires générés n’étant plus directement compatibles avec les standards comme HEVC, il faut donc adapter les logiciels et matériels des utilisateurs finaux. Il faut aussi considérer la complexité calculatoire élevée ajoutée au décodage.

Pour contre-balancer ces limites, des approches de simplification de texture avant encodage ont été proposées [455]. La nouvelle opération est effectuée uniquement coté codeur, et le flux de bits généré peut alors être décodé directement par le décodeur standard. Les approches de codage basé synthèse par prédiction des textures [291, 303] intègrent dans le boucle de codage le synthétiseur comme un nouveau mode prédiction, afin de réduire l'erreur résiduelle.

Nous avons proposé dans ce cadre méthodologique une synthèse de texture locale (appelée LTS) [4, 12, 49], qui prend place au sein de la boucle de codage hybride standard. Le principe est d'associer à chaque bloc un ensemble de blocs synthétiques respectant certaines contraintes statistiques, puis de laisser le codeur finaliser en sélectionnant le bloc synthétique minimisant le coût débit-distorsion. L'algorithme est décrit à la Figure 9.2. Il décrit comment est généré pour chaque bloc à coder  $B$  un ensemble de blocs synthétiques  $\underline{B}$  contenant  $n$  blocks ( $B_1, B_2, \dots, B_n$ ) qui sont visuellement similaires à  $B$  (car leur propre contexte  $Cx_{B_i}$  est proche de celui  $Cx'_B$ ). Notons que  $B$  lui-même appartient à  $\underline{B}$ . Un sous-ensemble  $\underline{B}'$  de  $\underline{B}$ , constitué de blocs correspondant encore mieux au contexte  $Cx'_B$  donné, peut-être construit. L'encodeur code ensuite le bloc courant  $B$  en le remplaçant successivement par chacun des blocs  $B_j$  de  $\underline{B}'$ , pour retenir à l'issue de cette étape le  $B_j$  produisant le débit-distorsion minimal. Ainsi, l'algorithme vise à encoder le bloc courant en remplaçant son contenu, par un autre visuellement similaire et moins coûteux à encoder. Le flux binaire généré est directement exploitable par le décodeur standard ciblé. Un exemple de résultat est illustré à la figure 9.3 où le décodeur est HEVC, un gain proche de 10% en bits a été obtenu avec l'encodeur LTS optimisé par rapport à celui standard.

Des approches différentes de codage existent qui tirent bénéfice des propriétés visuelles des textures en détectant celles non saillantes afin d'en tenir compte à l'encodage lors de l'optimisation débit-distorsion (c'est-à-dire en leur affectant moins de bits) [86, 96, 120, 156, 332, 412, 474]. Nous avons aussi produit des travaux dans ce cadre [4, 8, 10].

### 9.2.3 Métrique de similarité V1-E entre textures visuelles

Notre objectif a été alors de définir un modèle relatif à perception des textures aussi bien statiques que dynamiques, et donc apte à détecter leur degré de similarité visuelle. Ce modèle de similarité entre textures doit étendre ceux basés transformation et intégrer une modélisation plus poussée du SVH, sans être computationnellement trop complexe.

Nous avons vu au 9.2.1 que parmi les différentes étapes du traitement par le SVH, la zone corticale est considérée comme la plus active. Nous faisons aussi l'hypothèse que les textures ne sont définies ni par des formes, ni par des mouvements structurés<sup>1</sup>. Ceci conduit à exclure les niveaux supérieurs d'analyse associés à la perception des formes

---

1. Exactement nous faisons les hypothèses suivantes : les textures sont homogènes, le modèle est fovéal (aire spatiale inférieure à 4° d'angle visuel), la réponse oculaire est court terme (environ 200ms) et sans poursuite.



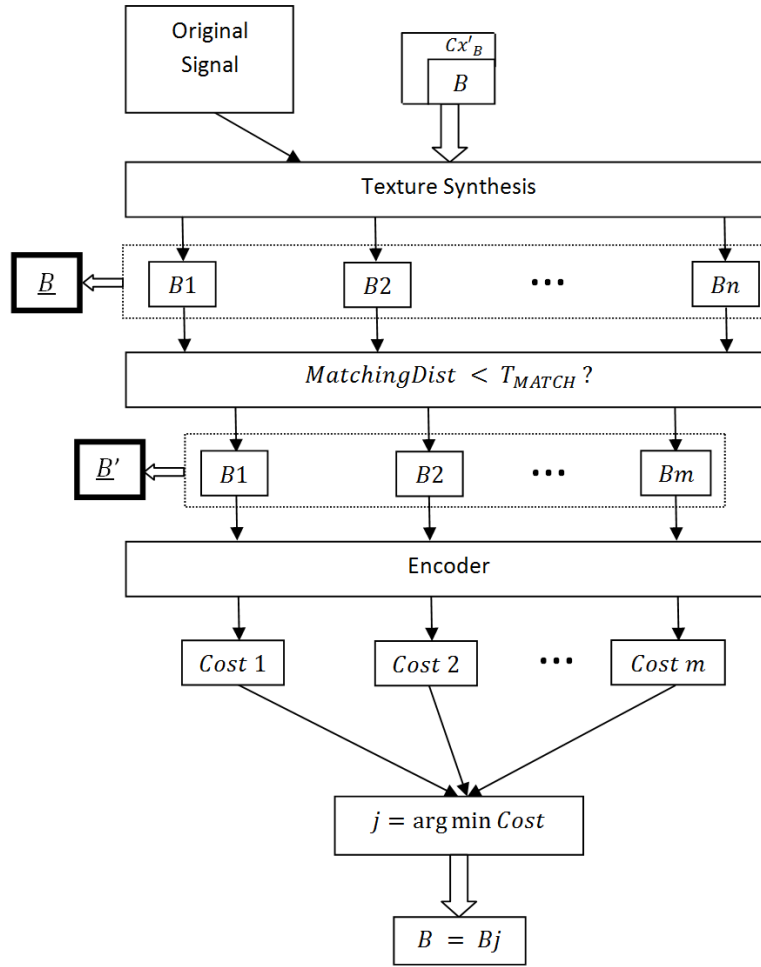


FIGURE 9.2 – Principe de notre approche de synthèse de texture locale [89].

(voie ventrale) et à la perception du mouvement (voie dorsale). Notre modèle ne considère donc que l'aire V1 (Fig. 9.1) où le traitement neuronal est généralement modélisé par un banc de filtres spatio-temporels passe-bandes. Nous avons nommé notre modèle V1-E (voir aussi la figure 9.4) car nous nous sommes appuyés sur les travaux décrits dans [92] qui considèrent comme caractéristiques les énergies des signaux des sous-bandes (d'où V1-E, « E » pour énergie). Ces énergies sont exploitées par notre métrique mesurant la similarité entre 2 textures  $T_1$  et  $T_2$ . D'abord les caractéristiques  $Ei_{T_1}$  et  $Ei_{T_2}$  de chaque  $i$ -ème sous-bande sont comparées (la formule est celle utilisée par la métrique STSIM [298]) :

$$Sim(Ei_{T_1}, Ei_{T_2}) = \frac{2Ei_{T_1}Ei_{T_2}}{Ei_{T_1}^2 + Ei_{T_2}^2 + \epsilon}$$

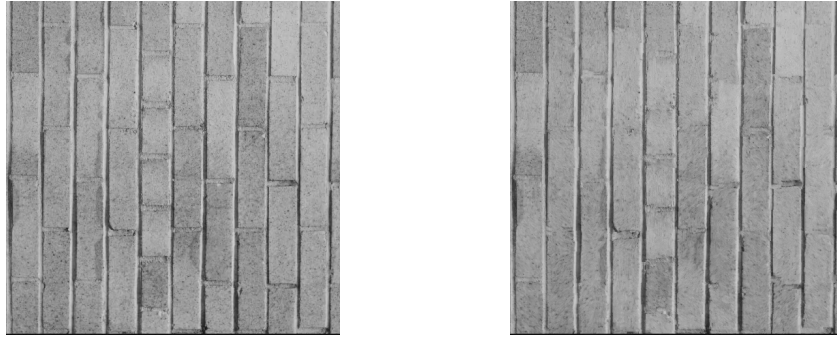


FIGURE 9.3 – Textures encodées/décodées avec HEVC (QP=27), à gauche : encodage standard , à droite : encodage LTS, le gain binaire est de 9.756% [89].

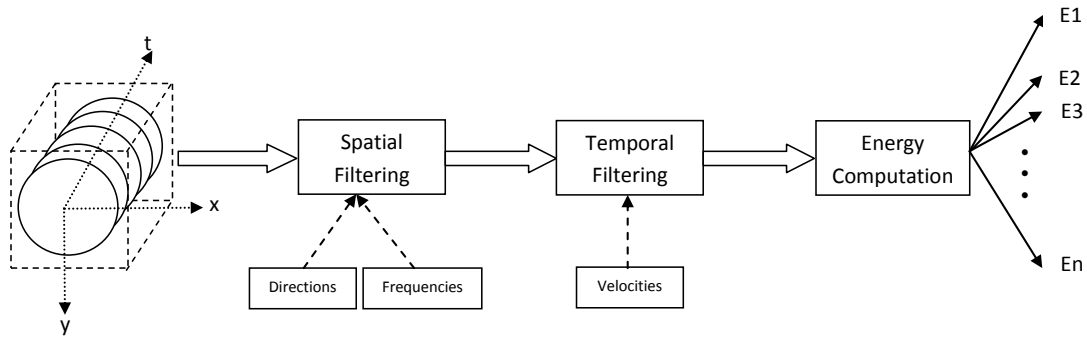


FIGURE 9.4 – Diagramme de bloc effectuant l'extraction des caractéristiques du modèle V1-E.

Le score de similarité globale est la moyenne considérant toutes les  $n$  sous-bandes :

$$SIM(T1, T2) = \frac{1}{n} \sum_{i=1}^n Sim(Ei_{T1}, Ei_{T2})$$

Nous avons testé notre métrique de similarité comme extracteur de caractéristiques pour une tâche de reconnaissance de texture [297] (c'est-à-dire pour une texture donnée, retrouver sa classe). Nous avons d'abord constitué une base (que nous avons nommé « HomoTex ») en capturant 38 vidéos de textures dynamiques, correspondant aux contraintes du modèle V1-E (c'est-à-dire un contenu dynamique homogène localisé spatialement et court terme), à partir des bases UCLA [378, 453], DynTex [239], et DynTex++ [229]. Chaque vidéo de 10s déterminant une classe de contenu, a été divisée en 50 échantillons d'environ 200ms chacun (des exemples sont donnés à la figure 9.5). La moitié des vidéos-échantillons a servi pour entraîner le modèle (calculs des énergies caractérisant la classe), l'autre moitié pour le tester. Les résultats expérimentaux montrent un taux de réussite

moyen de 99,84% au delà de celui à 95,51% de travaux LBT-TOP [336] de l'état de l'art.

#### 9.2.4 Application à la compression vidéo perceptuelle

La première étape pour lier V1-E à la compression vidéo, a été de mener une expérimentation psychophysique visant à mesurer la sensibilité visuelle aux distorsions de textures codées par le standard HEVC. L'intérêt est d'estimer, pour une texture donnée, la quantité d'information visuelle pouvant être éliminée sans provoquer de gêne visuelle (il s'agit donc d'information visuellement redondante). L'expérimentation a pour objectif d'estimer une fonction psychométrique [497, 504, 509] via une tâche de détection de seuil. Nous utilisons, parmi les méthodes adaptatives [499, 577, 642], celle UML « Updates Maximum Likelihood » [91, 162], et l'observateur est amené à choisir, parmi 2 vidéos codées à des débits différents  $R_1$  (ancrage) et  $R_2$ , laquelle il préfère (une capture écran du test est à la figure 9.6). Un résultat est donné à la figure 9.7, où le débit relatif est défini par  $R_r = (R_2 - R_1)/R_1$ , et le point important est celui dit d'égalité subjective pour lequel les vidéos comparées ont la même qualité perçue. Dans le cas de la figure 9.7, le point d'égalité subjective indique un débit relatif d'environ  $-10\%$ <sup>2</sup> (pour une qualité visuelle identique). Nous avons construit les courbes psychométriques de 25 textures issues de la base HomoTex (présentée à la section 9.2.3), chacune comprimée à 3 débits différents (donc 75 courbes en tout). Un résultat typique est illustré à la figure 9.8 validant l'idée que (pour la majorité des textures étudiées), bien que les points d'égalité subjective soient différents pour les trois QP, des gains en débit sont exploitables par le codeur. Nous avons ensuite vérifié si les énergies V1-E extraites permettaient de reconnaître une classe de texture donnée pour en déduire son taux de redondance visuelle (afin de choisir les bons paramètres de codage, le QP par exemple). Notre modèle de régression linéaire utilisant 9 énergies a montré de bonnes performances [10], néanmoins le procédé implique un traitement en amont du codeur, une opération directement au sein de la boucle de codage hybride serait préférable [86, 128, 172, 223, 236].

Nous nous sommes donc attachés à produire un codeur vidéo HEVC optimisé, car intégrant directement le modèle de similarité entre textures visuelles, et générant un flux binaire compatible avec le décodeur standard [86, 96]. Une première piste a été de modifier le schéma classique d'optimisation débit-distorsion afin qu'il devienne une optimisation débit-similarité<sup>3</sup>. Concrètement il s'agit de remplacer par des métriques perceptuelles exploitant des modèles de similarité entre textures visuelles, les métriques de distorsion utilisées par défaut par HEVC pour encoder un bloc [50] : c'est-à-dire celle SATD « Sum of Absolute Transformed Differences » employée afin de retenir un 1er lot des candidats les plus probables pour les modes de prédiction intra/inter, et de partition ; et celle SSD « Sum of Squared Differences » qui intervient ensuite afin de finaliser le choix. Nous avons produit des résultats expérimentaux préliminaires [4, 13,

2. Soit une baisse du débit de 10%.

3. Plus exactement il faudrait dire débit-dissimilarité.

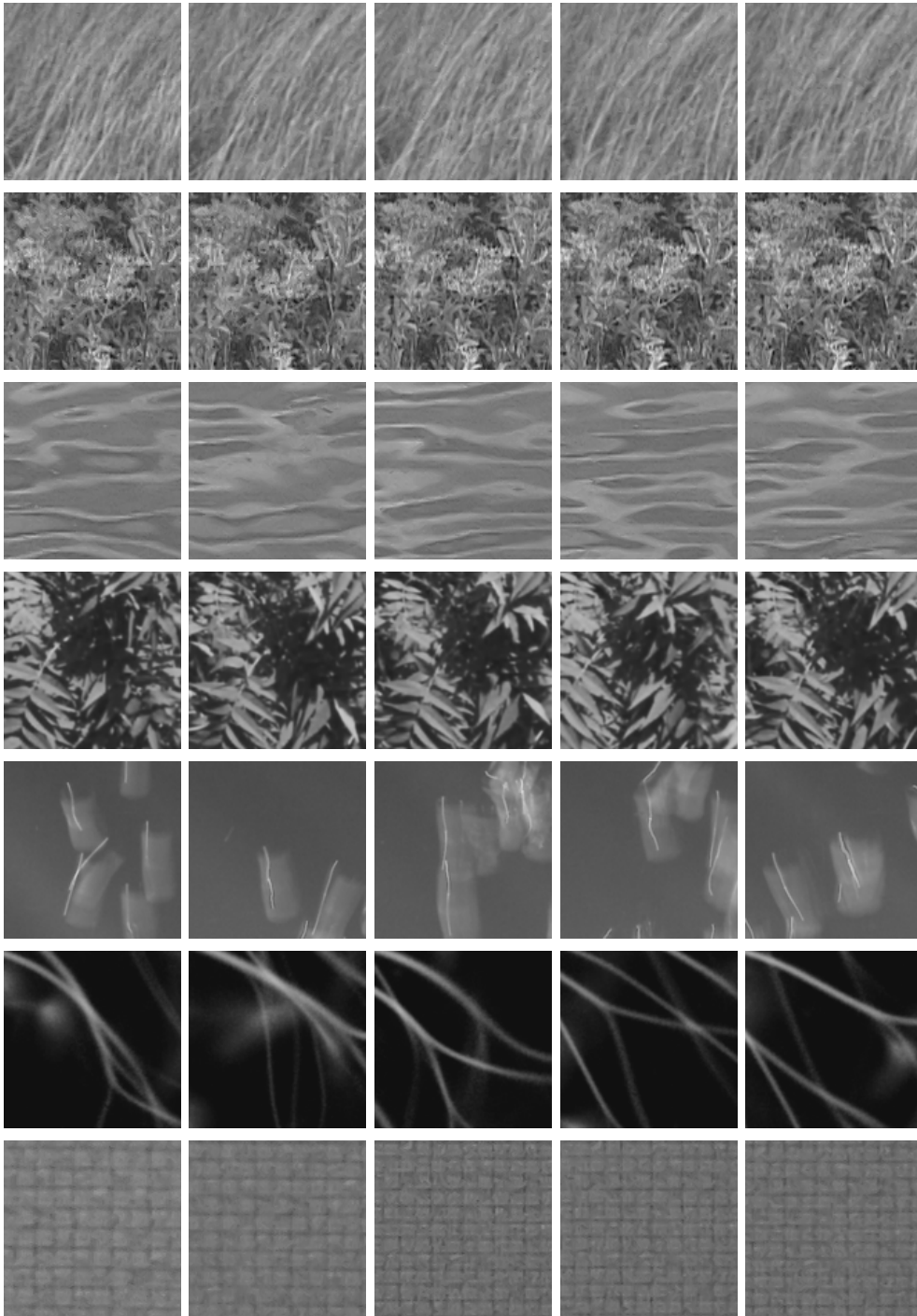


FIGURE 9.5 – 7 classes de textures issues de la base HomoTex utilisée lors des tests. Chaque ligne montre la 1ère image de 5 vidéos de textures dynamiques appartenant à une même classe donnée. Chaque vidéo a une résolution  $128 \times 128$  et dure typiquement  $200ms$ .



FIGURE 9.6 – Vue du logiciel utilisé pour le test de comparaison entre 2 vidéos de textures comprimées à des débits différents.

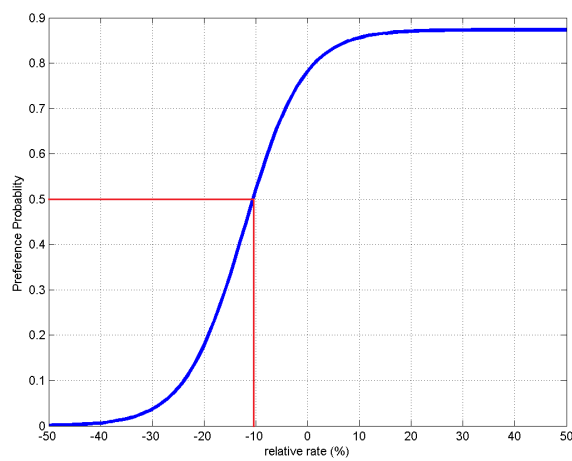


FIGURE 9.7 – Fonction psychométrique illustrant la probabilité de préférence des observateurs vis-à-vis des résultats de codage-décodage d'une texture donnée. Le point d'égalité subjective est repéré en rouge.

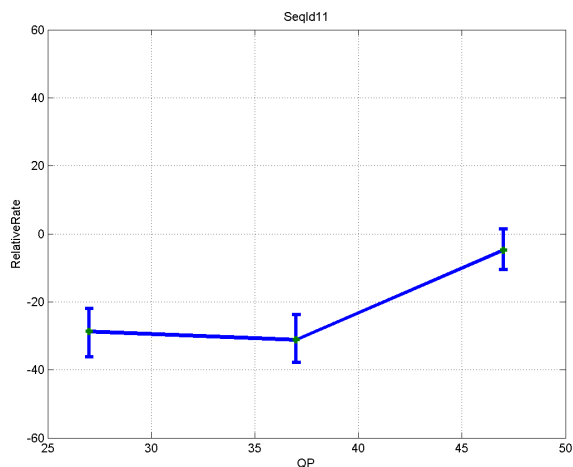


FIGURE 9.8 – Gains relatifs en débit, obtenus aux points d’égalité subjective pour une texture donnée et 3 débits (QP) différents. Les barres verticales indiquent les intervalles de confiance à 95%.

32, 14] relatifs au codage d’images fixes de textures<sup>4</sup> donc via le codeur HEVC en mode intra. Nous avons retenu les métriques de similarité STSIM [298] et, pour l’optimisation débit-distorsion, la distance de Gabor [561], ces 2 métriques sont conceptuellement très proches de celle V1-E. Une nette amélioration visuelle a été obtenue pour des niveaux de compression élevés et en particulier avec des textures au contenu structuré (voir un exemple à la figure 9.9). L’analyse a aussi montré la modification de stratégie du codeur optimisé retenant en moyenne (et indépendamment du QP) des blocs de prédiction plus larges, alors que le codeur standard les subdivise plus. Cependant une métrique comme STSIM n’est pas linéaire<sup>5</sup> ce qui la rend inopérable dans un schéma d’optimisation Lagrangienne.

Pour surmonter ce problème, nous avons proposé une solution faisant correspondre au résultat donné par la métrique classique MSE « Mean Squared Error » qui mesure la similarité entre pixels<sup>6</sup>, sa valeur équivalente dans le domaine perceptuel et exprimant la similarité entre textures visuelles. D’abord des expériences psychophysiques ont été conduites, afin d’établir ce lien entre différence calculée (c’est-à-dire la distorsion MSE entre la texture originale et celle codée par HEVC) et différence perçue. Nous avons adopté le test MLDS [282, 311, 462] « Maximum Likelihood Difference Scaling Methodology » et conduit l’observateur à choisir entre 2 paires de vidéos décodées donc

4. Les textures utilisées sont issues des bases Brodatz [51] et QualTex [230].

5. Une conséquence est qu’en cas de division d’un bloc, la somme des dissimilarités élémentaires des sous-blocs, n’est pas égale à la dissimilarité du bloc complet.

6. La métrique MSE (ou EQM « Erreur Quadratique Moyenne ») est proche de celle SSD utilisée par les codeurs standards, la différence est que SSD ne fait pas la moyenne des erreurs quadratiques mais les cumule.

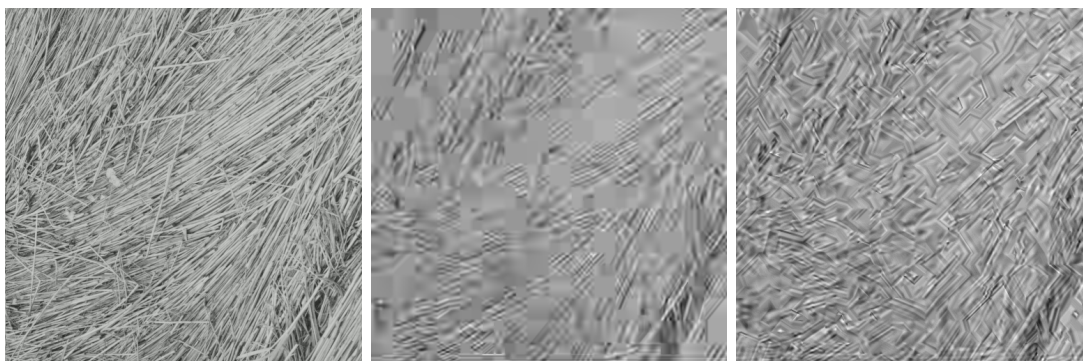


FIGURE 9.9 – Résultats obtenus en remplaçant les métriques SATD et SSD par celle STSIM, les codeurs HEVC utilisent un  $QP = 51$ , avec : à gauche l'image originale, au milieu celle encodée avec le codeur HEVC classique, et à droite celle encodée avec le codeur HEVC optimisé.

bruitées, laquelle paire présente le plus de différence visuelle (une capture écran du test est à la figure 9.10). Le test a été effectué sur un ensemble représentatif de vidéos de textures visuelles issues de la base HomoTex. Exactement les textures de la base ont été préalablement catégorisées en 8 classes, chacune reflétant un comportement débit-distorsion propre lorsque encodée par HEVC, les 8 textures les plus représentatives de ces classes ont donc été retenues pour le test MLDS (ces textures sont présentées à la figure 9.11). Les résultats de ce dernier ont montré que, vis-à-vis de la relation entre différences perçues et celles objectives, on peut distinguer globalement deux types de texture visuelle : l'un montrant un écart entre les deux différences, et l'autre ne présentant qu'une proportionnalité. La figure 9.12 illustre ces 2 comportements, on peut y remarquer que les fonctions de correspondance ont été linéarisées par morceaux. Afin de tirer profit des écarts de différences exploitables par le codeur, le MSE calculé sur un bloc de  $N$  pixels, est transformé en « SSD perçu ( $SSD_p$ ) » suivant :  $SSD_p = (\alpha MSE + \beta) \times N$ , où  $\alpha$  et  $\beta$  caractérisent le segment de droite considéré de la fonction de mise en correspondance propre à la texture. Le multiplicateur de Lagrange  $\lambda_p$ , paramètre clé de l'optimisation débit-« distorsion perçue » du codeur HEVC amélioré, s'exprime alors simplement :  $\lambda_p = -\frac{\partial SSD_p}{\partial r} = \left(\frac{\partial SSD_p}{\partial SSD}\right) \times \left(-\frac{\partial SSD}{\partial r}\right) = \alpha \times \lambda$ , où  $\lambda$  est celui du codeur standard. L'intérêt de l'approche a été vérifié via des tests subjectifs où les résultats obtenus avec le codeur HEVC standard<sup>7</sup> ont été comparés avec ceux du codeur HEVC amélioré intégrant notre métrique perceptuelle : après estimation des fonctions psychométriques en considérant différents niveaux de qualité et un ensemble de 24 textures en dehors de celles déjà retenues pour le test MLDS, les points d'égalité subjective ont montré une baisse moyenne du débit de 23,8% [4, 9, 11] (voir aussi la figure 9.13). Enfin, nous avons montré que les caractéristiques issues du modèle V1-E peuvent-être utilisées, pour prédire la classe débit-distorsion de la texture et donc sa fonction de mise en

7. Nous avons utilisé le software de référence HM 16.2 [50].



FIGURE 9.10 – Vue du logiciel utilisé pour le test MLDS de comparaison entre paires de vidéos codées.

correspondance entre distorsions. le modèle de prédiction basée SVM « Support Vector Machine » utilisait 8 énergies V1-E [8, 9].

### 9.3 Discussion

Ce travail a donc traité du concept de similarité visuelle entre textures. Nous avons passé en revue la plupart des études sur la perception des textures statiques et dynamiques, en ce concentrant sur celles qui conduisent au développement de modèles de similarité. Il a d'abord été nécessaire de définir et généraliser la notion de texture visuelle, puis nous avons proposé le modèle V1-E de similarité entre elles, V1-E se range alors dans la catégorie des modèles basés transformation en lien avec le traitement cortical du SVH. L'intérêt de V1-E a été étudié, nous l'avons préalablement testé avec succès comme extracteur de caractéristiques pour une tâche de reconnaissance de textures visuelles issues de la base HomoTex que nous avons assemblée, mais c'est surtout l'amélioration des systèmes de compression des images et des vidéos qui a motivé notre travail. Nous avons alors exploré différentes voies possibles en écartant celles produisant un flux comprimé incompatible avec les standards de compression :

- Nous avons initialement développé un procédé de synthèse de texture statique locale, qui procède uniquement coté codeur par remplacement du bloc de texture, si des contraintes statistiques et un gain débit-distorsion sont satisfaits. Le procédé a été breveté.





FIGURE 9.11 – Les 8 textures visuelles retenues pour le test MLDS.

- Nous avons ensuite utilisé V1-E comme extracteur de caractéristiques de la texture visuelle afin de prédire son taux de redondance visuelle exploitable par le codeur. Il s’agit donc d’une opération de catégorisation de la texture visuelle en amont du codage.
- Nous avons aussi cherché à intégrer notre procédé d’amélioration directement au sein de la boucle de codage hybride. D’abord une substitution de la métrique de distorsion standard par une mesure de la dissimilarité visuelle a été tentée. Cependant l’optimisation lagrangienne inhérente à l’encodage réclame des propriétés linéaires que n’ont pas les métriques proches conceptuellement de V1-E.
- L’ultime solution a été de mettre en oeuvre dans la boucle de codage, une fonction de mise en correspondance entre les mesures de distorsion objective standard et de distorsion perçue. Le procédé a montré une amélioration significative des performances débit-qualité de la norme de compression vidéo standard HEVC. Le modèle V1-E peut alors être utilisé comme extracteur de caractéristiques de la texture visuelle afin de prédire sa fonction de correspondance.

Ces travaux présentent une approche de codage des textures visuelles où la compatibilité directe avec le standard HEVC est très contraignante, car cela impose de s’intégrer au sein de la boucle de codage afin d’y piloter les décisions débit-distorsion prises à court terme. Par rapport à l’objectif plus général de codage vidéo basé contenu, l’approche la plus efficace serait certainement de combiner un ensemble de techniques supplémentaires. Nous en avons étudié certaines, par exemple la catégorisation des contenus permettant d’adapter les réglages d’encodage, ou encore la très efficace synthèse de textures. Cependant ces opérations impliquent une segmentation en amont du codeur (cela rejoint la stratégie de codage mise en avant au chapitre 7), et la description dans le flux comprimé du procédé de synthèse n’est pas prise en charge par les standards vidéo actuels.

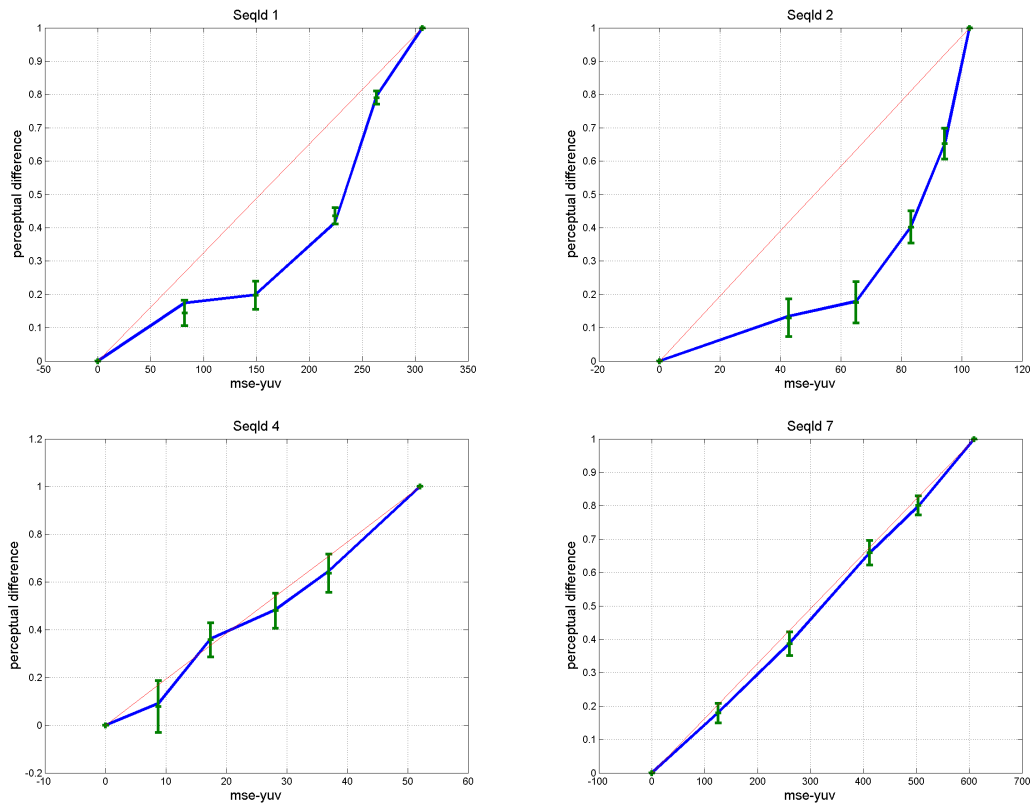


FIGURE 9.12 – Fonctions de correspondance, obtenues via les tests subjectifs MLDS, pour 4 textures visuelles : en haut un écart exploitable existe entre les distortions mesurées par le MSE (sur l'axe horizontal) et les distortions perçues (sur l'axe vertical) ; en bas il n'y a qu'une simple relation proportionnelle entre les distortions objectives et subjectives.

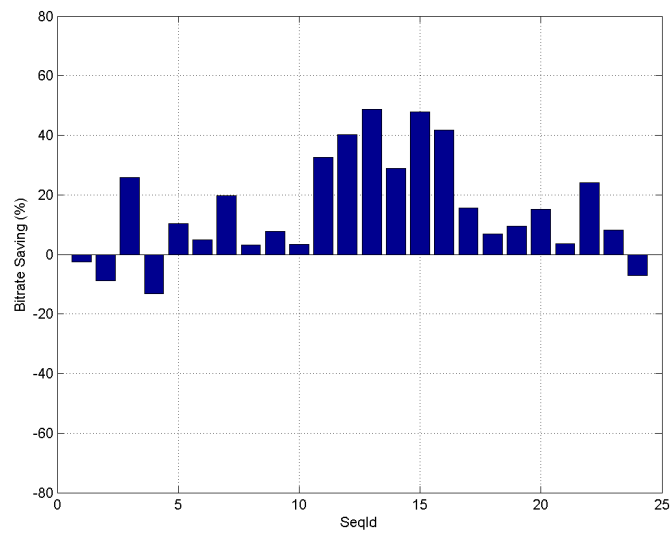


FIGURE 9.13 – Baisse du débit moyen pour 24 textures visuelles (chacune identifiée par un bâton).

## Chapitre 10

# Représentation efficace des contenus immersifs 3D

### 10.1 Contexte et problématique

Le développement des technologies permettant d'un côté de capturer le monde environnant avec ses multiples dimensions, et de l'autre de le présenter de façon immersive, explique la popularité croissante des services de vidéo augmentée, virtuelle ou mixte [135]. Entre les deux, demeure le besoin de représenter et de compresser les données volumétriques, afin de les distribuer ou de les stocker.

Ces données visuelles volumiques décrivent le monde réel via un nuage de points 3D, où ces derniers sont des voxels équivalents aux pixels des images 2D. Ces points sont caractérisés par leur géométrie (c'est-à-dire leur position et leur taille dans l'espace) et leurs attributs (c'est-à-dire leur couleur, leur réflectance, etc.). Lorsque la scène est dynamique, topologies et attributs évoluent en fonction du temps. Contrairement aux scènes 3D synthétiques où la structure des points est directement organisée afin de représenter chaque objet par un maillage polygonal (c'est-à-dire une liste de points et de leurs connexions afin de former les facettes et les bords), le nuage de points capturés n'est pas structuré. L'avantage du nuage de points est sa grande flexibilité notamment pour le rendu (si besoin graduable) des objets dynamiques [98]. Il représente de façon explicite la géométrie de l'objet 3D, ce qui est très utile pour combiner les rendus d'objets réels et virtuels, par exemple en réalité augmentée ou pour un service de communication immersive. Cependant, bien qu'un grand nombre de méthodes efficaces de codage des objets maillés ait été produit [255, 284, 534, 547, 548], il n'existe pas encore, début 2017, de méthodes efficaces pour les nuages de points capturés [57].

Avec Amira Filali et Nicolas Normand, nous sommes intéressés par ces problématiques de représentation des données volumiques, que nous voulons aborder en se mettant à l'intersection d'approches issues de traitement du signal et de la modélisation sous forme de graphes, et de la géométrie discrète. Nous pensons par exemple que, côté codeur,

des approches de quantification vectorielle algébriques et arborescentes seraient pertinentes pour structurer et représenter sur plusieurs niveaux la géométrie des nuages de points, et que coté décodeur, les méthodes d'interpolation des contenus discrets seraient efficaces pour améliorer le rendu des scènes 3D. Le schéma général de codage/décodage étudié est alors celui de la figure 10.1 où sont indiquées en gris les étapes ciblées par le travail de thèse.

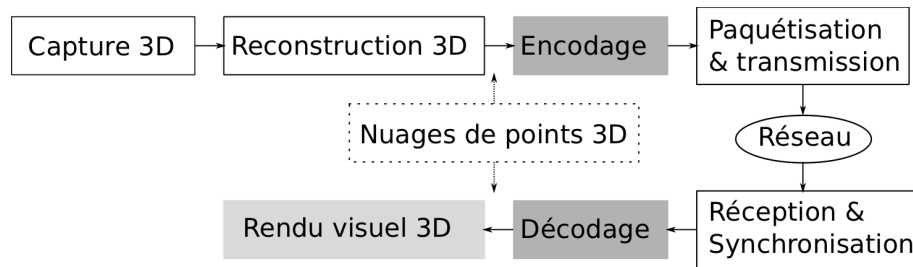


FIGURE 10.1 – Chaîne de codage/décodage de données immersives.

En parallèle, l'année 2017 est marquée par l'intérêt du groupe MPEG pour la compression des contenus immersifs capturés. Un calendrier est mis en place, avec l'organisation d'un appel d'offre portant sur le PCC pour « Point Cloud Compression », et l'objectif est de produire en 2020 un standard ISO. Exactement MPEG a identifié 3 cas d'usage [85], chacun correspondant à une catégorie de contenus 3D avec une approche respective de codage :

- le L-PCC « LIDAR Point Cloud Compression » pour les données acquises dynamiquement (catégorie 3) avec des systèmes de type LIDAR ;
- le S-PCC « Surface PCC » pour les données statiques (catégorie 1) avec haute densité de points ;
- le V-PCC « Video-based PCC » pour les contenus dynamiques (catégorie 2) exploités en réalité augmentée (AR).

Cet appel d'offre [62] a donc déjà abouti en octobre 2017, à l'issue d'un processus d'évaluation, à la production de 3 codeurs de référence appelés « test models » [52, 53] chacun ciblant respectivement l'une des 3 catégories de contenu. Le choix des technologies de codage a été effectué après comparaison de leurs performances (c'est-à-dire de la qualité des contenus 3D codés-décodés), ces dernières ayant été évaluées à partir de lots de données test selon des méthodes d'évaluation objectives (par les 2 métriques de qualité D1 « point-à-point » et D2 « point-à-plan » [68, 71]) et subjectives (via une méthodologie ACR [532]).

Parmi ces 3 schémas de codage, celui V-PCC nous intéresse moins, car il vise à projeter les données 3D dynamiques (profondeur et attributs) sur des images 2D+t afin de pouvoir les coder classiquement via un codec comme HEVC. Les technologies S-PCC

et L-PCC, qui vont certainement se combiner en une approche commune dite G-PCC « Geometry-based » [54], nous intéressent plus car il s’agit représenter le nuage de points, en décomposant l’espace 3D en une structure hiérarchique cubique obtenue typiquement par des découpages octree.

## 10.2 Travaux menés et premiers résultats

Nos travaux se sont donc rapidement orientés vers la représentation et le codage des données visuelles volumiques. Dans un premier temps, les nuages de points considérés sont statiques. L’état de l’art montre que les méthodes déjà proposées sont, soit dédiées aux contenus synthétiques [284], soit souffrent de faibles performances de compression spatiale et temporelle [67, 153] lorsqu’il s’agit de contenu naturel capturé. Pour les signaux issus de capteurs 3D, il faut réaliser une représentation efficace, graduable en résolution et en qualité, de la géométrie de la scène 3D. Tout en permettant le traitement (échantillonnage irrégulier, filtrage, transformation) des signaux caractéristiques des attributs photométriques afin de mieux les représenter et coder.

De nombreux travaux se sont intéressés à la compression de nuage de points essentiellement statiques. Souvent une décomposition octree de l’espace 3D est d’abord réalisée [200, 277, 360] afin de localiser les points et coder efficacement de façon entropique cette information. Les attributs (couleurs, normales) sont ensuite ajoutés relativement à l’arbre de décomposition. Différent types de représentations sont alors possibles : moyennes des attributs couleurs, coefficients DCT [575] ou d’ondelettes [59], etc. Les techniques récentes de traitement des signaux sur des graphes [106, 163], où les couleurs sont décomposées en les projetant sur les vecteurs propres de graphes Laplacien, offrent à ce niveau un ensemble d’outils pertinents. Le codage des attributs peut être prédictif [73, 77], et les approches sont particulièrement efficaces lorsque la structure hiérarchique arborescente est mise à profit pour obtenir des descriptions par approximations successives, une étape d’optimisation débit-distorsion lagrangienne peut-être ajoutée [73].

### 10.2.1 Approches de découpage

Nous proposons d’adapter l’approche de quantification vectorielle algébrique et arborescente (QVAA) [33, 28, 30] qui produit directement une description multi-niveaux de l’espace 3D, afin de représenter et coder la géométrie de la scène 3D capturée. Exactement la QVAA réunit un ensemble de techniques de quantification où le codage est opéré via un arbre de décision. Ses avantages sont une complexité calculatoire réduite grâce à l’utilisation de dictionnaires de représentants simples basés sur des réseaux réguliers de points (« lattices » en anglais) tronqués et emboîtés, et une structure arborescente permettant une représentation progressive s’adaptant à la répartition des points dans l’espace vectoriel de dimension 3 dans le cas du nuage de points. La QVAA utilise donc un ensemble hiérarchique de réseaux algébriques tronqués, construit de telle sorte qu’il soit possible d’emboîter un réseau tronqué ayant une échelle inférieure dans une cellule

de Voronoï du réseau à l'échelle immédiatement supérieure. Il faut donc définir un facteur d'échelle  $b$  entre les réseaux tronqués successifs de la hiérarchie (voir le principe à la figure 10.2). L'utilisation du réseau algébrique cubique permet un emboîtement optimal et une quantification rapide [113].

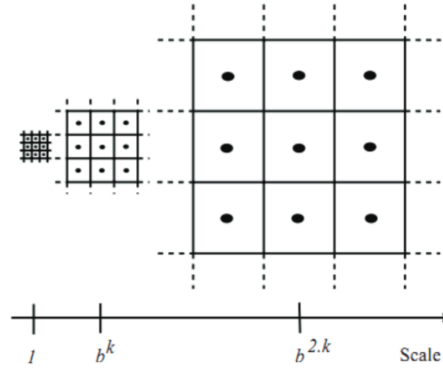


FIGURE 10.2 – Principe de la hiérarchie de réseaux tronqués s'emboîtant, ici pour le réseau cubique et en dimension 2. Lorsque le facteur d'échelle  $b = 3$ , cela produit le découpage d'un Voronoï en  $(3 \times 3)$  cellules [30].

Avec Amira nous innovons par rapport au schéma initial, en ajoutant la possibilité d'un découpage 3D octree  $(2 \times 2 \times 2)$  à celui  $(3 \times 3 \times 3)$ . L'objectif est d'offrir plus de liberté parmi les procédés de découpage de l'espace 3D afin de mieux s'adapter à la répartition spatiale du nuage de points. La nouvelle approche est décrite à la figure 10.3, le schéma à droite illustre la phase d'initialisation, et celui à gauche la quantification, avec :

- à l'étape (0) le point  $X$  (caractérisé par ses coordonnées) du nuage est centré et normalisé afin d'être projeté dans une boule au sein d'un premier cube unitaire (la racine de l'arbre).  $X_0$  représente la position moyenne de tous les points du nuage, et le facteur de normalisation est  $F = 1/(2 \times \sqrt{E_{max}})$  où  $\sqrt{E_{max}}$  est la distance euclidienne maximale entre un point et  $X_0$ .
- aux étapes (1) et (2), le point  $X$  au sein d'un cube unitaire donné est re-normalisé par  $b$  puis décalé de  $U$ . Ces 2 paramètres sont fonctions du découpage souhaité, pour celui  $(2 \times 2 \times 2)$  :  $b = 2$  et  $U = (1/2, 1/2, 1/2)$  ; pour celui  $(3 \times 3 \times 3)$  :  $b = 3$  et  $U = (0, 0, 0)$ .
- à l'étape (3), l'algorithme de quantification rapide [113] est mis en oeuvre afin de produire, par de simples opérations d'arrondi, les coordonnées du vecteur représentant  $Y_i$  caractéristique du nouveau cube auquel appartient  $X$ . Avec le jeu de paramètres choisi,  $Y_i = (y_1, y_2, y_3)$  avec  $y_i \in \{-1, 0, +1\}$  dans la cas du découpage  $(3 \times 3 \times 3)$ , et  $y_i \in \{0, +1\}$  dans la cas du découpage  $(2 \times 2 \times 2)$ ,  $Y_i$  est utilisé à l'étape (4) pour centrer le point puis effectuer un éventuel prochain niveau de quantification (c'est-à-dire un nouveau découpage de cube).

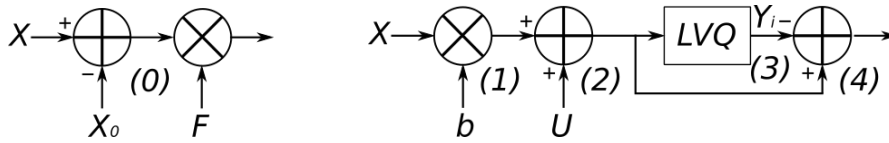


FIGURE 10.3 – Schémas de quantification avec, à gauche l'initialisation, et à droite la méthode de découpage ( $2 \times 2 \times 2$ ) ou ( $3 \times 3 \times 3$ ).

La figure 10.4 donne l'exemple de la construction d'un dictionnaire par découpage de l'espace 3D à partir du nuage de points initial « Teapot » [80].

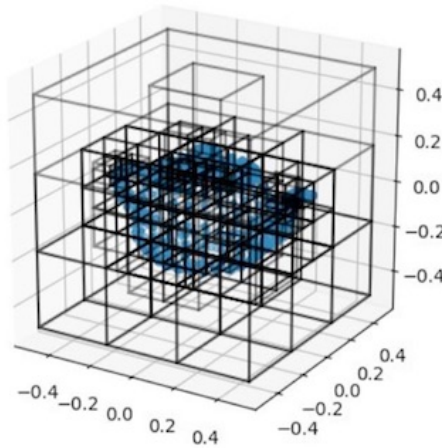


FIGURE 10.4 – Exemple d'un découpage 3D à partir du nuage de points « Teapot » [80].

### 10.2.2 Métriques et optimisation débit-distorsion

La structure arborescente est obtenue en associant à chaque nœud une cellule de Voronoï, dont les fils seront les cellules emboîtées après découpage. La racine de l'arbre représente le cube initial n'ayant pas encore été découpé et contenant le nuage de points en entier. Pour construire l'arbre représentant le nuage de points et donc découper l'espace, une approche de programmation dynamique est utilisée qui fait croître de façon progressive la structure arborescente de telle sorte qu'à l'issue de chaque boucle un nœud terminal, donc une feuille, soit découpée [600]. Le processus itératif réalise simultanément la description progressive de la structure topologique du nuage de points, et la simplification de ce dernier pour le coder efficacement. Cette simplification est opérée en associant un représentant aux points du nuage d'une feuille, et le représentant le plus simple est certainement le centre du cube (c'est-à-dire le vecteur représentant  $Y_i$  de la figure 10.3). Un des objectifs lors de la construction de l'arbre est donc de réduire globalement la distorsion entre les points du nuage et le dictionnaire de ses représentants. En parallèle, le processus itératif de construction de l'arbre occasionne une augmentation progressive du coût binaire de sa représentation car il faut coder la liste des nœuds



découpés et des représentants associés aux feuilles.

Finalement le codage du nuage de points se formule comme un problème d'optimisation débit-distorsion classique où il s'agit de minimiser la distorsion globale  $D(s_j)$  associée à l'arbre  $s_j$  obtenu à la  $j$ -ième boucle du processus itératif, avec la contrainte que le débit global  $R(s_j)$  demeure inférieur ou égal au débit cible  $R_T$  :

$$\min_{s_j} D(s_j) \text{ tel que } R(s_j) \leq R_T$$

Ce problème peut être résolu par une approche lagrangienne [546] et il s'agit de minimiser conjointement :

$$\min_{s_j} L(s_j) \text{ où } L(s_j) = D(s_j) + \lambda \times R(s_j)$$

$\lambda$  est le multiplicateur de Lagrange qui contrôle le compromis débit-distorsion. Nous adoptons alors l'algorithme BFOS [599] où l'optimisation est opérée localement en associant un retour marginal à chaque nœud découpé  $n_i$  de l'arbre :  $\lambda(n_i) = \Delta D / \Delta R$ , où  $\Delta D$  et  $\Delta R$  indiquent respectivement la baisse de distorsion et la hausse de débit occasionnées. Donc à chaque boucle du processus de construction de l'arbre, le nœud découpé à conserver est donc celui offrant  $\lambda(n_i)$  maximal.

Nous estimons le coût en bits du découpage d'un nouveau nœud (c'est-à-dire son ajout dans l'arbre), en considérant un encodage entropique. Pour évaluer objectivement la distorsion locale au sein d'une cellule  $C_i$ , entre les points  $X$  du nuage et leur représentant  $Y_i$ , nous utilisons les 2 métriques  $D1$  et  $D2$  adoptées par MPEG [71] :

— celle « point-à-point » où  $\|\overrightarrow{XY_i}\|$  est une norme euclidienne :

$$D_{p2p}(X, Y_i) = \sum_{X \in C_i} \|\overrightarrow{XY_i}\| \quad (10.1)$$

— celle « point-à-plan » où le produit scalaire est calculé entre  $\overrightarrow{XY_i}$  et  $\overrightarrow{N_X}$  qui est la normale à la surface de l'objet en  $X$  (si besoin, cette normale peut-être estimée [79]) :

$$D_{p2plan}(X, Y_i) = \sum_{X \in C_i} |\overrightarrow{XY_i} \cdot \overrightarrow{N_X}| \quad (10.2)$$

### 10.2.3 Premiers résultats

Pour les résultats expérimentaux nos résultats de codage sont comparés à ceux du codeur de référence MPEG G-PCC [54], et nous avons utilisé les 3 nuages de points statiques « Soldier », « Long Dress » et « Loot », tous issus des données du MPEG-3DG [53]. Nous avons comparé différentes stratégies possibles pour le découpage d'un nœud de l'arbre : pour un premier codeur seul le découpage ( $2 \times 2 \times 2$ ) est utilisé, pour un second seul celui ( $3 \times 3 \times 3$ ) est employé, enfin pour le troisième cas dit « hybride »,

l'un ou l'autre des deux découpages est possible et retenu selon le critère d'optimalité lagrangienne. Nous avons aussi évalué l'apport du choix de la méthode de calcul de la distorsion locale au noeud, soit en la calculant suivant la métrique « point-à-point » (voir l'équation eq. 10.1), ou soit suivant celle « point-à-plan » (voir l'équation eq. 10.2).

A des fins de comparaison, nous avons configuré les algorithmes afin que les nuages de points décodés comptent environ le même nombre d'éléments (20000 points). Les performances présentées dans le tableau 10.1 sont mesurées en termes de distance de Hausdorff symétrique [63] où la mesure utilisée est une erreur quadratique moyenne « point-à-point ».

L'analyse des résultats montre que dans un seul cas (« Loot » avec la distorsion locale « point-à-point ») le mode  $(3 \times 3 \times 3)$  se révèle supérieure aux autres. Cela peut s'expliquer car ce mode opère un découpage brusque et donc moins précis de l'espace 3D. Le mode  $(2 \times 2 \times 2)$  permet de mieux contrôler le découpage progressif de l'espace, il produit pour « Soldier » et « Long Dress » et avec la distorsion locale « point-à-point » les meilleurs résultats.

Le mode « Hybride », lorsque la métrique « point-à-point » est utilisée, apparait logiquement comme un compromis entre les 2 autres modes de découpage. Dans les cas « Loot » et de « Long Dress », les résultats objectifs sont aussi supérieurs à ceux obtenus avec le codeur de référence.

Les résultats lorsque le calcul de la distorsion locale est basé « point-à-plan » sont toujours en retraits de ceux obtenus lorsqu'il est basé « point-à-point » (sauf dans 2 cas : « Soldier » avec le découpage hybride, et « Loot » pour le mode  $(2 \times 2 \times 2)$ ). De plus, le gros écart obtenu selon le choix du calcul de la distorsion locale avec « Long Dress » pour un découpage  $(2 \times 2 \times 2)$  peut sembler surprenant. Pour expliquer ces effets, il faut regarder subjectivement les rendus des nuages de points décodés, comme ceux présentés à la figure 10.5 et obtenus directement à l'aide la librairie Open3D [58]. Pour ces images, les attributs couleur ont été ajoutés : la couleur d'un représentant  $Y_i$  étant la moyenne des attributs des points au sein du cube  $C_i$ .

Nuages	MPEG G-PCC	Notre codeur avec découpage :		
		$(3 \times 3 \times 3)$	$(2 \times 2 \times 2)$	Hybride
<i>Soldier</i>	16.48	18.72 <b>19.02</b>	11.2816 <b>14.32</b>	18.70 <b>14.45</b>
<i>Long Dress</i>	16.53	18.74 <b>19.89</b>	9.28 <b>61.08</b>	15.89 <b>17.74</b>
<i>Loot</i>	16.54	15.33 <b>17.65</b>	18.77 <b>10.49</b>	15.36 <b>19.13</b>

TABLE 10.1 – Comparaison des schémas de codage en termes de distance de Hausdorff symétrique. Dans le cas de notre codeur, le chiffre en gras est obtenu lorsque le calcul de la distorsion locale est basée « point-à-plan », sinon elle est basée « point-à-point ».

Tout d'abord on peut remarquer à la figure 10.5(b) la pauvre qualité du nuage de

points rendu où aucun détail n’apparaît, car le nuage a été codé avec l’algorithme de référence MPEG trop fortement contraint en débit. Dans le cas de notre méthode, l’optimisation débit-distorsion vise à découper plus grossièrement les zones de l’espace ayant un contenu simple (donc à leur allouer moins de points), de façon à pouvoir découper plus finement les zones complexes. Ainsi pour un nombre d’environ 20000 points, les nuages de points des figure 10.5(c-f) codés-décodés selon notre méthode, présentent une qualité de rendu acceptable, il est d’ailleurs difficile à ce niveau de distinguer l’avantage objectif du découpage ( $2 \times 2 \times 2$ ) de la figure 10.5(c) sur celui ( $3 \times 3 \times 3$ ) de la figure 10.5(d). Il faut comparer les rendus des figures 10.5(e) et 10.5(f) pour comprendre l’influence du choix de la métrique locale de distorsion. Dans le cas avec celle « point-à-plan » (figure 10.5(e)), nous pouvons remarquer que l’objet rendu comporte beaucoup plus de « trous » que dans le cas avec la métrique locale « point-à-point » (figure 10.5(f)). Ces « trous » sont dus à la réduction du nombre de points au niveau des zones les plus lisses de l’objet, certains détails peuvent alors être mieux rendus (par exemple l’extrémité de l’arme). Néanmoins ces « trous » sont très pénalisants lorsque les performances sont mesurées objectivement via la distance de Hausdorff symétrique. Enfin la figure 10.5(g) montre une exploitation possible de la propriété multi-échelle de notre approche : par rapport au rendu de la figure 10.5(e), les « trous » ont été comblés en ajoutant plus de points (le nuage compte alors environ 30000 points), et le rendu est alors quasi identique à celui original de la figure 10.5(a).

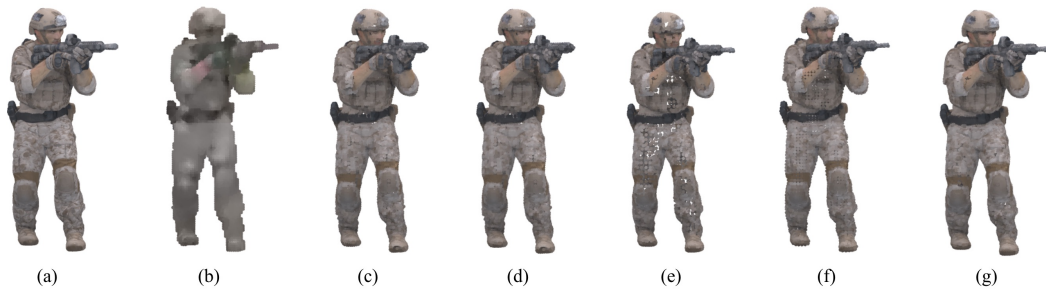


FIGURE 10.5 – Rendus du nuage de points « Soldier ». (a) Original comptant 1089091 points, (b) Nuage codé/décodé avec MPEG PCC pour  $\approx 20000$  points, (c-f) Nuages codés/décodés via notre approche, pour  $\approx 20000$  points, et différents modes de découpage (le calcul de la distorsion locale est « point-à-point » sauf pour (e)) : (c) ( $2 \times 2 \times 2$ ), (d) ( $3 \times 3 \times 3$ ), (e) Hybride et distorsion locale « point-à-plan », (f) Hybride. (g) Découpage hybride et distorsion locale « point-à-plan », pour  $\approx 30000$  points.

### 10.3 Futurs travaux

Nous avons donc proposé un schéma de codage de nuages de points 3D statiques issus de données capturées immersives. Le schéma de codage avec pertes s’appuie sur une approche basée QVAA. Notre méthode innove en combinant les deux schémas de

découpage  $(2 \times 2 \times 2)$  et  $(3 \times 3 \times 3)$  des cellules de Voronoï cubiques, et en utilisant les deux métriques « point-à-point » et « point-à-plan » pour estimer la distorsion géométrique locale. L'arbre représentant à plusieurs échelles la topologie 3D du nuage est construit de manière itérative et en fonction d'un critère d'optimisation débit-distorsion. Comparés au codeur de référence MPEG PCC, les résultats expérimentaux montrent des performances objectives et subjectives prometteuses.

Dans nos travaux futurs, nous projetons d'améliorer ce codeur en travaillant suivant de multiples axes :

- Le choix du représentant associé à la feuille :  
Le schéma actuel utilise un simple point (le centre du cube), il serait judicieux de pouvoir aussi utiliser des primitives géométriques plus complexes (facettes, points multiples) adaptées à certaines zones, et y associer les attributs (couleur, normales, coefficients obtenus après transformation sur les graphes) afin d'aboutir à un processus de représentation plus précis et sémantiquement significatif.
- Le calcul du débit réel lié au codage de la géométrie de la scène et de ses attributs :  
Une estimation précise du coût de découpage d'un noeud est nécessaire afin au niveau de l'optimisation débit-distorsion. La méthode de codage sans perte de l'arbre représentant la géométrie de la scène est bien maîtrisée lorsqu'il s'agit d'un découpage octree  $(2 \times 2 \times 2)$  [57]. Elle repose schématiquement sur l'utilisation d'un code sur un octet pour décrire la subdivision du noeud, et du codage arithmétique du flux de ces codes lorsqu'on parcourt l'arbre. Il faudra donc adapter ce schéma au cas du découpage hybride  $((2 \times 2 \times 2)$  versus  $(3 \times 3 \times 3))$  qui est plus complexe car le mode de découpage choisit n'est plus implicite. Il faudra également considérer le codage des attributs. Les méthodes actuelles reposent sur le codage arithmétique de la liste des attributs associés aux feuilles de l'arbre. Des approches plus sophistiquées mettent en oeuvre une approche LOD « Level of Detail » afin de générer le nuage de points sur plusieurs niveaux de raffinement, en contraignant les distances entre les points décodés : les points d'un niveau inférieur étant plus éloignés spatialement que les points du niveau juste supérieur. Nous pensons que notre structure hiérarchique arborescente pourrait être utilisée pour effectuer une description LOD en exploitant alors les représentants associés aux noeuds intermédiaires. Il faudrait aussi considérer leur codage arithmétique, et aussi gérer des contraintes géométriques supplémentaires lors de la construction de l'arbre.
- Le calcul de la distorsion locale et de celle globale :  
Une estimation précise de la distorsion locale associée à un noeud est nécessaire au niveau de l'optimisation débit-distorsion. Au vu de nos premiers résultats expérimentaux retenant les règles d'évaluation objective de MPEG, l'usage de la métrique « point-à-point » l'emporte sur celle « point-à-plan ». Cependant les résultats subjectives montrent également que la distribution spatiale des points est judicieusement répartie car il y en a plus dans les zones chargées en détails. Il serait

donc intéressant de considérer, à partir des points décodés, l'ajout d'une opération de régularisation des surfaces [60] afin d'améliorer le rendu visuel 3D final.

— L'élagage de l'arbre :

Notre approche de construction de l'arbre représentant la géométrie du nuage procède itérativement par découpage local de l'espace 3D. L'intérêt est de faire croître progressivement la structure et d'ajuster au plus juste les besoins en mémoire et calcul. Il serait intéressant, en utilisant le même critère d'optimisation débit-distorsion, de comparer les résultats précédents à ceux obtenus par une approche d'élagage d'un arbre complet de très grande taille (car représentant très finement la scène). Ceci produit en théorie une solution globale optimale [609]

— La représentation et le codage de nuage de points dynamiques :

Il s'agit de coder le nuage de points en y ajoutant la dimension temporelle. Quelques travaux existent déjà, par exemple dans [153] un codec basé octree exploite via des opérations XOR les redondances temporelles existantes entre les flux des arbres décrivant la scène 3D en mouvement. Dans [67], une extension est ajoutée pour le codage des attributs via un codage d'image classique. Thanou et al. ont décrit dans [84] un algorithme capable de prédire pour mieux les coder, les différences entre les structures octree relatives aux images adjacentes du nuage dynamique. Le codage avec pertes est basé transformation sur graphe et utilise une décomposition sur ondelettes spectrales. Le codage des attributs couleur est proposé dans [106] à base de transformation locale sur des sous-graphes autour des points de la structure topologique précédente.

Dans notre cas, il s'agit donc d'étendre les fonctionnalités de notre schéma afin de l'adapter efficacement à la représentation et au codage des nuages de points dynamiques. Nous pensons notamment que, pour une catégorie de scène capturée donné, un modèle de représentation géométrique arborescente pourra être obtenu en scindant l'arbre en deux éléments complémentaires [27] avec : une racine commune à tous les nuages de la catégorie, et de branches à ajouter afin de composer finement le nuage d'une scène 3D donnée. L'identification des catégories de scène et la construction de leur racine respective pourra se faire en tirant profit de techniques avancées d'apprentissage, et pour une scène donnée, le codage des branches et des attributs relatifs aux nuages adjacents temporellement se basera sur des approches prédictives.

## Chapitre 11

# Conclusion générale et perspectives

Ce mémoire a donc retracé mon itinéraire scientifique depuis 2002 (ou plus exactement 2004), avec mes contributions au codage vidéo perceptuel menées au sein des équipes IVC/IRCCyN puis IPI/LS2N qui m'ont offert un cadre particulièrement propice à cette recherche. Nous avons vu que mon travail, suivant l'évolution générale qui s'est imposée pendant la période aux chercheurs des disciplines STIC, s'est développé au sein de projets de recherche, et dans mon cas il y a eu successivement les projets ANR ArchiPEG (cf. chapitre 3), PERSEE (chapitre 4), puis l'ITN PROVISION (cf. chapitre 5)<sup>1</sup>. Le modèle de la recherche projet organisée via des consortiums de partenaires complémentaires, se révèle en effet souvent nécessaire pour aborder la compétition internationale en codage vidéo. Néanmoins il faut aussi désormais surveiller l'impact de la fondation AOMedia « Alliance for Open Media »<sup>2</sup>, créée en septembre 2015 par des compagnies « poids lourds de l'Internet », et qui vise au développement de formats vidéo ouverts et libres (un exemple est la spécification AV1 publiée en 2018 et relative à la diffusion de flux vidéos sur l'Internet). L'AOMedia appuie donc des projets alternatifs aux standards ITU-T/ISO<sup>3</sup>.

Mes activités ont été de deux types, l'un d'ordre organisationnel avec des participations actives aux projets de recherche, ma responsabilité la plus grande ayant été la coordination scientifique du projet PERSEE (sachant que j'ai eu en parallèle des responsabilités pédagogiques importantes à la tête du département Informatique de Polytech'Nantes). L'autre type d'activité est lié à la supervision de thèses et de masters. Nos contributions au codage vidéo perceptuel, c'est-à-dire concrètement la conception d'un ensemble d'outils et de méthodes visant à améliorer la qualité perceptuelle des vidéos codées-décodées, ont été à ce niveau appréhendées de deux façons. L'ambition a d'abord été de redéfinir intégralement un schéma de codage à partir d'une représentation de l'in-

---

1. Et actuellement je participe au projet ANR DISSOCIE qui a débuté en 2018.

2. Le site web de l'« Alliance for Open Media » : [aomedia.org](http://aomedia.org)

3. Un exemple est le projet Daala dont le site web est : [xiph.org/daala](http://xiph.org/daala)

formation conforme au modèle perceptuel. Nous l'avons expérimenté lors de la thèse de Guillaume Jeannic (cf. chapitre 6) pour la compression des images fixes. Cependant il est difficile de produire un codeur complet capable de concurrencer les performances de standards parfaitement optimisés et fruits de décennies de travaux. L'objectif est plutôt devenu celui d'accompagner, pour l'orienter et l'améliorer, le développement des nouveaux standards de compression vidéo et/ou le déploiement des services associés. Nous l'avons mené dans le cadre des travaux thèses qui ont suivi, avec celle d'Olivier Brouard (cf. chapitre 7) définissant un outil de pré-analyse avant encodage en flux H.264/AVC optimisé, la thèse de Junle Wang (cf. chapitre 8) développant des modèles d'attention visuelle 2D&3D, puis celle de Karam Naser (cf. chapitre 9) exploitant la similarité entre textures visuelles pour améliorer HEVC, enfin il y a la thèse en cours d'Amira Filali (cf. chapitre 10) qui propose la quantification vectorielle de nuages de points dans le cadre du futur MPEG-I PCCV-PCC.

### Perspectives

Nous l'avons montré au travers ce mémoire : la recherche en codage vidéo est fortement orientée pour répondre aux appels régulièrement lancés par les organismes de standardisation ITU-T et ISO (c'est-à-dire actuellement le JVET), appels qui eux-mêmes répondent aux besoins liés aux avancées technologiques (par exemple dans les domaines de la capture et de l'affichage des vidéos numériques, des réseaux, des calculateurs, etc.) et au développement de nouveaux services. Nous l'avons aussi expliqué au chapitre 10, la communauté du codage vidéo est mobilisée actuellement pour élaborer (et produire en 2020) le futur standard MPEG-I relatif aux formats immersifs. Parmi les propositions en gestation, nous pouvons identifier deux approches de codage fondamentalement différentes. L'une, appelée V-PCC « Video-based PCC », vise à projeter les données 3D volumiques (profondeur et attributs) sur des surfaces  $2D + t$  afin d'obtenir des vidéos dont le compression sera assurée par une boucle de codage prédictif hybride. Il s'agit donc de tirer profit de techniques qui ont fait leur preuve mais qui demeurent finalement de conception classique. L'autre, appelée G-PCC « Geometry based PCC », conserve les caractéristiques intrinsèques des données discrètes dynamiques 3D, et procède via un paradigme de codage différent en représentant et comprimant la géométrie de la scène puis ses attributs. C'est dans ce cadre que nous allons développer à courts termes nos recherches, en se positionnant à l'intersection d'approches issues de la quantification vectorielle, du filtrage sur graphe, et de l'interpolation géométrique discrète. Le détail des travaux prévus a été exposé dans la conclusion 10.3 du chapitre 10.

Mes projets de recherche à moyens et longs termes sont définitivement liés au traitement du signal appliqué à la représentation et au codage des images et des vidéos numériques. On peut déjà remarquer que tous les outils et méthodes de codage perceptuel présentés dans ce mémoire, offrent des pistes d'amélioration toujours en phase avec les besoins du codage vidéo, aussi bien celui obéissant au paradigme de codage classique que celui retenant le nouveau.

Il faut en parallèle tenir compte du potentiel énorme de progrès que peuvent apporter

les approches d'apprentissage profond liées au « machine learning », elles ont déjà fait leur preuve en vision par ordinateur [78, 87, 143], mais jusqu'à présent elles ont eu peu d'application au codage : des travaux ont été réalisés en image fixe pour améliorer les transformations en se basant sur des architectures de type « auto-encodeur » [55, 61, 72, 74], d'autres en vidéo afin d'optimiser la qualité de reconstruction coté décodeur par interpolation/extrapolation de l'information entre images avec des réseaux de neurones profonds [65, 69]. Le « machine learning » serait efficace par exemple lorsqu'il s'agit de finaliser une stratégie de codage basé contenu, sachant que la conception des bases de données correspondantes intégrant les futurs formats très haute résolution (par exemple, le 8K HDR à 120 images par seconde) mobilisent aussi la communauté [56, 81]. Des pistes d'amélioration possibles, en retenant celles qui me motivent, sont :

- Le développement de modèle de calcul de la distorsion perçue est toujours nécessaire. Les métriques de qualité doivent être compatibles avec la contrainte de linéarité requise pour l'optimisation lagrangienne débit-distorsion, et implantables facilement car ce dernier processus intervient lors de la prise de décision à court terme des codeurs hybrides. Nous avons proposé une solution dans le cadre du travail de thèse de Karam Naser au paragraphe 9.2.4 via une fonction de mise en correspondance adaptée au type de la texture visuelle considérée, le « machine learning » permettrait de généraliser à une gamme complète de types de contenu. Afin d'obtenir des modèles perceptuels plus précis, il faudrait aussi élargir en tenant compte d'un ensemble plus large de caractéristiques (mouvement, couleur, fréquence d'images, résolution, etc.) [66, 131, 524].
- Le développement de métriques mesurant la fidélité sémantique (on parle aussi de métriques à long terme) offre d'autres axes de recherche contribuant à améliorer la qualité des contenus décodés. Dans ce cas l'analyse en amont de la vidéo est nécessaire afin de segmenter le contenu, et de retenir un codage cohérent pendant la durée de vie des objets extraits. C'est la stratégie de codage que nous avons déjà explorée lors de la thèse d'Olivier Brouard présentée au chapitre 7. Les méthodes d'apprentissage permettraient d'améliorer le procédé en renforçant l'exploitation des cohérences spatio-temporelles entre segments, et l'intégration de modèles perceptuels prenant mieux en compte le mouvement [593].  
Ces approches ouvrent aussi concrètement le champ au prometteur codage vidéo basé contenu évoqué au paragraphe 5.1, car après segmentation et classification du contenu, il est devenu possible de personnaliser le codage de chaque objet identifié. Un projet de recherche passionnant serait de construire un tel codeur afin de mettre en œuvre en les combinant les approches de codage perceptuel que nous avons développés et testés de façon unitaire (par exemple les transformations adaptées de la partie 6, ou le codage des textures de la partie 9), et aussi les nouveaux outils liés à la prédiction, l'interpolation, etc.
- Nous avons aussi vu au paragraphe 10.2.2 relatif au travail de thèse conduit avec Amira Filali, que le besoin en métrique de qualité demeure lorsqu'il s'agit de coder



suivant le nouveau paradigme les nuages de points dynamiques. La classification des types de contenu dans ce contexte et l'optimisation du rapport débit-qualité, sont des pistes de recherche qui m'intéressent.

Au delà du codage vidéo, nous pensons que les travaux que nous menons sur la structuration de l'information géométrique de nuages de points 3D peuvent avoir des applications en vision (par exemple par rapport à des problématiques de guidage d'un robot mobile, de cartographie 3D ou de localisation) [70, 85]. Nous conduisons aussi d'autres travaux prometteurs en géométrie discrète et transformation Mojette, où le remplacement du lattice<sup>4</sup> cubique par des lattices plus compacts apportent des améliorations dans la précision de représentations des données [7], le développement d'applications en analyse d'images (médicales par exemple) ou en transmission reste à faire.

---

4. Un « lattice » est un réseau régulier de points.

# Table des figures

2.1	Boucle d'encodage hybride type pour HEVC (extrait de [164]). . . . .	23
2.2	Vue chronologique des principaux évènements décrits dans le mémoire et en particulier ceux relatifs à mes contributions au codage perceptuel. . . .	26
3.1	Différentes résolutions existantes en SD et HD TV (extrait de [315]). . . .	31
3.2	La plateforme matérielle ArchiPEG (extrait de [304]). . . . .	33
3.3	Architecture générale de la plate-forme matérielle ArchiPEG avec son système d'interconnexion à double anneaux (extrait de [304]). . . . .	34
4.1	L'organisation en tâches de PERSEE avec leurs contributeurs, le partenaire responsable de la tâche est indiqué en rouge. . . . .	39
4.2	Inpainting : principe des approches de type $K$ -NN (d'après [95]). . . . .	41
4.3	Protocole expérimental pour l'évaluation des séquences FVV synthétisées (d'après [110]). . . . .	43
5.1	Le modèle de codage vidéo basé contenu du projet PROVISION (d'après [151]).	47
5.2	Interdépendance des WP1 à 5 (c'est-à-dire des lots de travaux à caractère scientifique) de PROVISION (d'après [150]). . . . .	49
5.3	Le pilotage de PROVISION (d'après [151]). . . . .	50
6.1	Le schéma de codage utilisé (d'après la présentation de soutenance de thèse [278] disponible sur HAL). . . . .	55
6.2	Les étapes de la technique de lifting pour un signal 1D (d'après [278]). . .	56
6.3	Exemple, d'un schéma lifting basé colonne, avec la prédiction d'un échantillon d'une colonne impaire (en gris) à partir de divers échantillons de colonnes paires (en blanc) (d'après [278]). . . . .	57
6.4	Pour les 3 implantations testées, les exemples de l'étape de prédiction du schéma lifting basé colonne, pour une orientation $(p,q)$ de filtrage d'environ $33^\circ$ par rapport à l'horizontal (d'après [278]). . . . .	58

6.5	Comparaison de l'implantation de notre TOO par rapport à une plus classique (d'après la présentation de soutenance de thèse [278] disponible sur HAL).	59
6.6	Les étapes de notre TOO (d'après [23]).	60
6.7	Schéma de codage/décodage des contours extraits utilisé dans un contexte de TOO (d'après [278]).	66
6.8	Approche multi-résolution pour le codage arithmétique de contours. A : prochain code différentiel de Freeman à coder/décoder. B : les N derniers codes différentiels de Freeman codés/décodés. C : écart entre la direction orthogonale au gradient à la résolution inférieure et la direction donnée par les 2 derniers pixels codés/décodés. (d'après [22]).	67
6.9	Exemple d'une CSF montrant que l'œil humain est plus sensible aux variations d'amplitude d'un signal moyenne fréquence que haute fréquence (d'après [278]).	67
6.10	À gauche : la décomposition en TO 2D en sous-bandes représentée dans un espace spatio-fréquentiel. À droite : la décomposition en ondelettes dans l'espace fréquentiel (d'après [278]).	68
6.11	Capture d'écran de l'application permettant de modifier les pas de quantification et les pondérations structurelles, tout en observant les images codées/décodées (d'après [278]).	68
7.1	Les étapes de la recherche exhaustive des prédictions intra et inter du logiciel de référence de H.264/AVC (extrait de [217]).	70
7.2	Architecture biologiquement plausible d'un modèle d'attention visuelle (extrait de [613]).	72
7.3	Pré-analyse de la vidéo en amont du codeur, avant traitement la vidéo a été découpée en plans homogènes (extrait de [217]).	73
7.4	Le cœur de l'outil de pré-analyse vidéo (extrait de [217]).	73
7.5	Tube spatio-temporel (extrait de [40]).	74
7.6	Estimation par accumulation des paramètres du mouvement global d'un segment temporel (extrait de [217]).	75
7.7	Indices de confiance des vecteurs mouvement (extrait de [40]).	76
7.8	Exemple de compensation du mouvement global de la séquence vidéo <i>New mobile and Calendar</i> (extrait de [40]).	77
7.9	Segmentation de l'espace d'accumulation des vecteurs mouvement non compensés (extrait de [217]).	78
7.10	Traitement Intra/Inter d'un segment temporel par approche markovienne (extrait de [217]).	79

7.11	Cartes de segmentation pour <i>Tractor</i> (segments 13 à 16) : segmentation basée mouvement (ligne du milieu) et approche markovienne (ligne du bas) (extrait de [217]). . . . .	81
7.12	Représentation des couleurs dans l'espace conique HSV (ou espace « Teinte, Saturation, Intensité ») (extrait de [38]). . . . .	82
7.13	Saillance temporelle fonction de la vitesse relative (extrait de [38]). . . . .	84
7.14	Cartes de saillance pour <i>Tractor</i> (segments 13 à 16), avec de haut en bas : les images originales, les cartes de saillance spatiale, celles de saillance temporelle, puis celles de saillance spatio-temporelle (extrait de [38]). . . . .	85
7.15	Classification des segments temporels en fonction des indices d'activité spatiale <i>IS</i> et temporelle <i>IT</i> (extrait de [217]). . . . .	86
7.16	Fonction de modulation du pas de quantification du macrobloc en fonction de son indice de saillance, $S^{SP-T}$ est la valeur moyenne des indices de l'image à laquelle appartient le macrobloc <i>i</i> (d'après la présentation de soutenance de thèse [217] disponible sur HAL). . . . .	88
8.1	Exemples de CDF utilisées, avec de gauche à droite : l'image originale (issue de la base LIVE [402]), puis les 3 CDF (de Sydney, Delft, puis Nantes) (extrait de [171]). . . . .	99
8.2	Exemples de cartes issues des expériences, avec de gauche à droite : l'image originale, la carte d'importance et celle de saillance (extrait de [242]). . . . .	100
8.3	Nombre de fixations oculaires sur (a) les sujets principaux, (b) les objets secondaires, et (c) les objets d'arrière-plan, pendant la durée d'observation (extrait de [242]). . . . .	100
8.4	Composition des stimuli : (a) arrière-plan (un bruit blanc), (b) positions des projections des objets sur l'écran, (c) distribution des objets en profondeur (extrait de [3]). . . . .	101
8.5	Schéma d'ensemble du modèle d'attention visuelle 3D proposé (extrait de [2]). . . . .	102
8.6	Exemples de cartes de saillance obtenues, avec de gauche à droite : (a) l'image originale, (b) la CDF (vérité de terrain), puis les cartes de saillance calculées selon (c) le modèle proposé (saillance de profondeur uniquement), (d) Itti, (e) Bruce, et (f) Hou (extrait de [171]). . . . .	103
8.7	Illustration du conflit entre accommodation et vergence. A gauche le cas de visualisation d'un contenu 2D : les yeux s'accommodent et convergent naturellement vers le même objet sur l'écran. A droite le cas de visualisation d'un contenu 3D stéréoscopique : les yeux s'accommodent sur l'écran mais convergent vers un objet ici en avant plan, les distances focale et de convergence différent, en dehors d'une plage de tolérance, inconfort visuel et fatigue visuelle apparaissent (extrait de [171]). . . . .	104

8.8	Exemples de stimuli : avec à gauche le cas d'un « SB-stimulus » dont le fond est net, et à droite le cas d'un « BB-stimulus » dont le fond a été flouté (extrait de [17]). . . . .	104
8.9	Construction des stimuli : les plans de profondeur bleu contiennent le papillon en avant du plan rouge qui est le fond. A un « BB-stimulus » (à gauche), sont associés sept ou huit « SB-stimuli » (à droite) (extrait de [17]). . . . .	105
9.1	Modèle hiérarchique du système visuel, d'après [101]. . . . .	108
9.2	Principe de notre approche de synthèse de texture locale [89]. . . . .	110
9.3	Textures encodées/décodées avec HEVC (QP=27), à gauche : encodage standard , à droite : encodage LTS, le gain binaire est de 9.756% [89]. . . . .	111
9.4	Diagramme de bloc effectuant l'extraction des caractéristiques du modèle V1-E. . . . .	111
9.5	7 classes de textures issues de la base HomoTex utilisée lors des tests. Chaque ligne montre la 1ère image de 5 vidéos de textures dynamiques appartenant à une même classe donnée. Chaque vidéo a une résolution $128 \times 128$ et dure typiquement $200ms$ . . . . .	113
9.6	Vue du logiciel utilisé pour le test de comparaison entre 2 vidéos de textures comprimées à des débits différents. . . . .	114
9.7	Fonction psychométrique illustrant la probabilité de préférence des observateurs vis-à-vis des résultats de codage-décodage d'une texture donnée. Le point d'égalité subjective est repéré en rouge. . . . .	114
9.8	Gains relatifs en débit, obtenus aux points d'égalité subjective pour une texture donnée et 3 débits (QP) différents. Les barres verticales indiquent les intervalles de confiance à 95%. . . . .	115
9.9	Résultats obtenus en remplaçant les métriques SATD et SSD par celle STSIM, les codeurs HEVC utilisent un $QP = 51$ , avec : à gauche l'image originale, au milieu celle encodée avec le codeur HEVC classique, et à droite celle encodée avec le codeur HEVC optimisé. . . . .	116
9.10	Vue du logiciel utilisé pour le test MLDS de comparaison entre paires de vidéos codées. . . . .	117
9.11	Les 8 textures visuelles retenues pour le test MLDS. . . . .	118
9.12	Fonctions de correspondance, obtenues via les tests subjectifs MLDS, pour 4 textures visuelles : en haut un écart exploitable existe entre les distorsions mesurées par le MSE (sur l'axe horizontal) et les distorsions perçues (sur l'axe vertical) ; en bas il n'y a qu'une simple relation proportionnelle entre les distorsions objectives et subjectives. . . . .	119
9.13	Baisse du débit moyen pour 24 textures visuelles (chacune identifiée par un bâton). . . . .	120

10.1 Chaîne de codage/décodage de données immersives. . . . . 122

10.2 Principe de la hiérarchie de réseaux tronqués s'emboîtant, ici pour le réseau cubique et en dimension 2. Lorsque le facteur d'échelle  $b = 3$ , cela produit le découpage d'un Voronoï en  $(3 \times 3)$  cellules [30]. . . . . 124

10.3 Schémas de quantification avec, à gauche l'initialisation, et à droite la méthode de découpage  $(2 \times 2 \times 2)$  ou  $(3 \times 3 \times 3)$ . . . . . 125

10.4 Exemple d'un découpage 3D à partir du nuage de points « Teapot » [80]. 125

10.5 Rendus du nuage de points « Soldier ». (a) Original comptant 1089091 points, (b) Nuage codé/décodé avec MPEG PCC pour  $\approx 20000$  points, (c-f) Nuages codés/décodés via notre approche, pour  $\approx 20000$  points, et différents modes de découpage (le calcul de la distorsion locale est « point-à-point » sauf pour (e)) : (c)  $(2 \times 2 \times 2)$ , (d)  $(3 \times 3 \times 3)$ , (e) Hybride et distorsion locale « point-à-plan », (f) Hybride. (g) Découpage hybride et distorsion locale « point-à-plan », pour  $\approx 30000$  points. . . . . 128

# Index

- IT*, 85, 86
- ST*, 85
- $S^{SP-T}$ , 84, 87
- 3D-HEVC, 22
- 3DNR, 38
- 3DTV, 37
- 3DVC, 37
  
- AccepTV, 34
- ACR, 88
- AMT, 65
- ANR, 28
- AOMedia, 131
- ArchiPEG, iv, 30
- AROC, 91, 95
- AV, 71, 81, 90
  
- BBC, 48
  
- CCLP, 91, 95
- CDF, 90, 95
- Codage vidéo basé contenu, 47
- CONTINT, 38
- CSD, 38
- CSF, 63
  
- Daala, 65
- DCT, 65
- DST, 65
  
- ENVIVIO, 30
- EQM, 20, 115
- ER, 47
- ESR, 47
- ETP, 38
  
- FP7, 46
- FVV, 37
  
- G-PCC, 123, 132
- GOP, v, 85
- Groupe Multimédia LTCI, 40
  
- H.264/AVC, iv, v, 21, 30–33, 69, 132
- HD, iv, 22
- HDTV, 37
- HEVC, iv, v, 22, 106, 132
- HFOrtho, 58
- HFReg, 58
- HFRes, 58
- HHI, 48
- HM, 22
- HomoTex, 111, 116
- HTM, 22
  
- IETR, 34, 37
- IMG (IETR), 34, 40
- INRIA Rennes, 37
- IPI, iv, 1, 6, 20
- IRCCyN, iv, 6, 20, 37
- IS, 86
- ITN, 46
- IVC, iv, 6, 20, 32, 39, 48
  
- JCT-VC, 21
- JM, 22, 69
- JPEG 2000, 54
- JVET, 22, 132
- JVT, 21
  
- KTA, 22, 65

L-PCC, 122  
LDI, 42  
LDV, 42  
LS2N, iv, 1, 20  
LTCL, 37  
LTS, 109

MDDT, 65  
MLDS, 115  
MOS, 88  
MPEG, 21, 122  
MPEG-I, 22  
MPEG-I PCC, 132  
MSCA, 46  
MSE, 115  
MV-HEVC, 22  
MVC, iv, 22, 37  
MVD, 42

PCC, 22, 122  
PERSEE, iv, 38, 131  
Programme BLANC, 38  
PROVISION, iv, 47  
PSNR, 20

QoE, 98  
QVAA, 123, 128

RIAM, 30  
RWTH, 48

S-PCC, 122  
SATD, 112  
Service Europe et recherche (LUNAM), 46  
SHVC, 22  
SSD, 112  
SVC, 22  
SVH, v, 21, 72, 96, 107  
SVM, 117  
SVT, 75  
SynDEx, 34

TEMICS, 39  
TO, 54  
TOO, 59

UE, 46  
UHD, 22  
UML, 112  
UoB, 48  
UoN, 48

V-PCC, 122, 132  
V1-E, 110  
VCEG, 21  
VITEC, 30, 38, 48  
VQEG, 23  
VVC, 22

WP, 48

x264, 86–88



# Bibliographie

- [50] C ROSEWARNE, B BROSS, M NACCARI, K SHARMAN et GJ SULLIVAN. *High Efficiency Video Coding (HEVC) Test Model 16 (HM 16)*, Joint Collaborative Team on Video Coding (JCTVC) of ITU-T SG16 WP3 and ISO. Rapp. tech. IEC JTC1/SC29/WG11, Document JCTVC-P1002.
- [51] USC-SIPI Dataset. URL : <http://sipi.usc.edu/database/>.
- [37] Amira FILALI, Vincent RICORDEL et Nicolas NORMAND. « Tree-Structured Point-Lattice Vector Quantization for 3-D Point Cloud Geometry Compression ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Nov. 2018.
- [52] 3D GRAPHICS. « PCC Test Model Cat13 ». In : *ISO/IEC JTC1/SC29/WG11 MPEG N 17762* (juil. 2018).
- [53] 3D GRAPHICS. « PCC Test Model Cat2 ». In : *ISO/IEC JTC1/SC29/WG11 MPEG N 17767* (juil. 2018).
- [54] 3D GRAPHICS. « PCC WD G-PCC ». In : *ISO/IEC JTC1/SC29/WG11 MPEG N 17771* (juil. 2018).
- [55] Nick JOHNSTON, Damien VINCENT, David MINNEN, Michele COVELL, Saurabh SINGH, Troy CHINEN, Sung JIN HWANG, Joel SHOR et George TODERICI. « Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks ». In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, p. 4385–4393.
- [1] Vassilios KRASSANAKIS, Matthieu Perreira Da SILVA et Vincent RICORDEL. « Monitoring Human Visual Behavior during the Observation of Unmanned Aerial Vehicles (UAVs) Videos ». In : *Drones* 2.36 (2018).
- [56] Alex MACKIN, Mariana AFONSO, Fan ZHANG et David BULL. « A study of subjective video quality at various spatial resolutions ». In : *IEEE International Conference on Image Processing (ICIP)*. 2018, p. 2830–2834.
- [57] Sebastian SCHWARZ, Marius PREDÄ, Vittorio BARONCINI, Madhukar BUDAGAVI, Pablo CESAR, Philip A CHOU, Robert A COHEN, Maja KRIVOKUCA, Sébastien LASSERRE, Zhu LI et al. « Emerging MPEG standards for point cloud compression ». In : *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* (2018).

- 
- [58] Qian-Yi ZHOU, Jaesik PARK et Vladlen KOLTUN. « Open3D : A Modern Library for 3D Data Processing ». In : *arXiv :1801.09847* (2018).
- [59] Siheng CHEN, Dong TIAN, Chen FENG, Anthony VETRO et Jelena KOVACEVIC. « Fast resampling of 3d point clouds via graphs ». In : *arXiv preprint arXiv :1702.06397* (2017).
- [60] David COEURJOLLY, Pierre GUETH et Jacques-Olivier LACHAUD. « Digital surface regularization by normal vector field alignment ». In : *International Conference on Discrete Geometry for Computer Imagery (DGCI)*. Springer. 2017, p. 197–209.
- [61] Michele COVELL, Nick JOHNSTON, David MINNEN, Sung Jin HWANG, Joel SHOR, Saurabh SINGH, Damien VINCENT et George TODERICI. « Target-quality image compression with recurrent, convolutional neural networks ». In : *arXiv preprint arXiv :1705.06687* (2017).
- [62] 3D GRAPHICS. « Call for proposals for point cloud compression v2 ». In : *ISO/IEC JTC1/SC29/WG11 MPEG N 17332* (avr. 2017).
- [63] Alireza JAVAHERI, Catarina BRITES, Fernando PEREIRA et João ASCENSO. « Subjective and objective quality evaluation of 3D point cloud denoising algorithms ». In : *International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE. 2017, p. 1–6.
- [64] Li LIU, Paul FIEGUTH, Yulan GUO, Xiaogang WANG et Matti PIETIKÄINEN. « Local binary features for texture classification : Taxonomy and experimental study ». In : *Pattern Recognition* 62 (2017), p. 135–160.
- [65] Ziwei LIU, Raymond A YEH, Xiaoou TANG, Yiming LIU et Aseem AGARWALA. « Video frame synthesis using deep voxel flow ». In : *Proceedings of the IEEE International Conference on Computer Vision*. 2017, p. 4463–4471.
- [66] Alex MACKIN, Katy C NOLAND et David R BULL. « High frame rates and the visibility of motion artifacts ». In : *SMPTE Motion Imaging Journal* 126.5 (2017), p. 41–51.
- [67] Rufael MEKURIA, Kees BLOM et Pablo CESAR. « Design, implementation, and evaluation of a point cloud codec for tele-immersive video ». In : *IEEE Transactions on Circuits and Systems for Video Technology* 27.4 (2017), p. 828–842.
- [68] Rufael MEKURIA, Sebastien LASERRE et Christian TULVAN. « Performance assessment of point cloud compression ». In : *Visual Communications and Image Processing (VCIP)*. IEEE. 2017, p. 1–4.
- [4] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Perceptual Texture Similarity for Machine Intelligence Applications ». In : *Visual Content Indexing and Retrieval with Psycho-Visual Models*. Sous la dir. de Jenny BENOIS-PINEAU et Patrick Le CALLET. Springer, 2017.

- [69] Simon NIKLAUS, Long MAI et Feng LIU. « Video frame interpolation via adaptive separable convolution ». In : *Proceedings of the IEEE International Conference on Computer Vision*. 2017, p. 261–270.
- [31] Vincent RICORDEL, Nicolas NORMAND et Jeanpierre GUÉDON. « Analysis of Mojette Transform Projections using Z3 lattice versus A3 lattice ». In : *Sino-French Workshop on Research Collaborations in Informations and Communication Technologies (SIFWICT)*. Qingdao, China. Juin 2017.
- [7] Vincent RICORDEL, Nicolas NORMAND et Jeanpierre GUÉDON. « Mojette Transform on Densest Lattices in 2D and 3D ». In : *International Conference on Discrete Geometry for Computer Imagery (DGCI)*. Sous la dir. de SPRINGER. Lecture Notes in Computer Science 10502. Vienne, Austria. Sept. 2017, p. 159–170.
- [70] Kazuo SUGIMOTO, Robert A COHEN, Dong TIAN et Anthony VETRO. « Trends in efficient representation of 3D point clouds ». In : *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE. 2017, p. 364–369.
- [71] Dong TIAN, Hideaki OCHIMIZU, Chen FENG, Robert COHEN et Anthony VETRO. « Geometric distortion metrics for point cloud compression ». In : *International Conference on Image Processing (ICIP)*. IEEE. 2017, p. 3460–3464.
- [72] George TODERICI, Damien VINCENT, Nick JOHNSTON, Sung JIN HWANG, David MINNEN, Joel SHOR et Michele COVELL. « Full resolution image compression with recurrent neural networks ». In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, p. 5306–5314.
- [73] Yiqun XU, Shanshe WANG, Xinfeng ZHANG, Shiqi WANG, Nan ZHANG, Siwei MA et Wen GAO. « Rate-distortion optimized scan for point cloud color compression ». In : *Visual Communications and Image Processing (VCIP)*. IEEE. 2017, p. 1–4.
- [74] Johannes BALLÉ, Valero LAPARRA et Eero P SIMONCELLI. « End-to-end optimization of nonlinear transform codes for perceptual quality ». In : *2016 Picture Coding Symposium (PCS)*. IEEE. 2016, p. 1–5.
- [75] Manuela CHESSA, Silvio P SABATINI et Fabio SOLARI. « A systematic analysis of a V1–MT neural model for motion estimation ». In : *Neurocomputing* 173 (2016), p. 1811–1823.
- [76] Olena CHUBACH, Patrick GARUS et Mathias WIEN. « Motion-Based Analysis and Synthesis of Dynamic Textures ». In : *Proc. of International Picture Coding Symposium PCS '16*. Nuremberg, Germany : IEEE, Piscataway, déc. 2016.
- [77] Robert A COHEN, Dong TIAN et Anthony VETRO. « Point cloud attribute compression using 3-D intra prediction and shape-adaptive transforms ». In : *Data Compression Conference (DCC)*. IEEE. 2016, p. 141–150.
- [78] Ian GOODFELLOW, Yoshua BENGIO et Aaron COURVILLE. *Deep learning*. MIT press, 2016.

- 
- [8] Chengyue MA, Karam NASER, Vincent RICORDEL, Patrick Le CALLET et Chunmei QING. « An adaptive Lagrange multiplier determination method for dynamic texture in HEVC ». In : *IEEE International Conference on Consumer Electronics-China (ICCE-China)*. Guangzhou, China. Déc. 2016, p. 1–4.
- [9] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « A foveated short term distortion model for perceptually optimized dynamic textures compression in HEVC ». In : *Picture Coding Symposium (PCS)*. Nuremberg, Germany. Déc. 2016.
- [10] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Estimation of Perceptual Redundancies of HEVC Encoded Dynamic Textures ». In : *International Conference on Quality of Multimedia Experience (QoMEX)*. Lisbon, Portugal. Juin 2016.
- [11] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Modeling the Perceptual Distortion of Dynamic Textures and Its Application in HEVC ». In : *IEEE International Conference on Image Processing (ICIP)*. Phoenix, Arizona, United States. Sept. 2016.
- [79] Przemyslaw PODULKA. « Selection of reference plane by the least square fitting methods ». In : *Advances in Science and Technology Research Journal* 10.30 (2016).
- [80] The Utah Teapot - CHM REVOLUTION. *Computer History Museum*. 2016.
- [81] German ROS, Laura SELLART, Joanna MATERZYNSKA, David VAZQUEZ et Antonio M LOPEZ. « The synthia dataset : A large collection of synthetic images for semantic segmentation of urban scenes ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, p. 3234–3243.
- [82] U. S. THAKUR et B. RAY. « Image coding using parametric texture synthesis ». In : *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*. Sept. 2016, p. 1–6. DOI : [10.1109/MMSP.2016.7813339](https://doi.org/10.1109/MMSP.2016.7813339).
- [83] Uday THAKUR, Karam NASER et Mathias WIEN. « Dynamic Texture Synthesis using Linear Phase Shift Interpolation ». In : *Proc. of International Picture Coding Symposium PCS '16*. Nuremberg, Germany : IEEE, Piscataway, déc. 2016.
- [84] Dorina THANOU, Philip A CHOU et Pascal FROSSARD. « Graph-based compression of dynamic 3D point cloud sequences ». In : *IEEE Transactions on Image Processing* 25.4 (2016), p. 1765–1778.
- [85] Christian TULVAN, R MEKURIA, Z LI et S LASERRE. « Use cases for point cloud compression ». In : *ISO/IEC JTC1/SC29 WG11 MPEG Output Document* (2016).
- [86] Long XU, Weisi LIN, Lin MA, Yongbing ZHANG, Yuming FANG, King N NGAN, Songnan LI et Yihua YAN. « Free-Energy Principle Inspired Video Quality Metric and Its Use in Video Coding ». In : (2016).

- 
- [87] Yann LECUN, Yoshua BENGIO et Geoffrey HINTON. *Deep learning*. T. 521. 7553. Nature Publishing Group, 2015, p. 436.
- [88] NV Kartheek MEDATHATI, Manuela CHESSA, Guillaume MASSON, Pierre KORNPROBST et Fabio SOLARI. « Decoding MT Motion Response for Optical Flow Estimation : An Experimental Evaluation ». Thèse de doct. INRIA Sophia-Antipolis, France ; University of Genoa, Genoa, Italy ; INT la Timone, Marseille, France ; INRIA, 2015.
- [12] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Local Texture Synthesis : A Static Texture Coding Algorithm Fully Compatible With HEVC ». In : *International Conference on Systems Signals and Image Processing (IWSSIP)*. London, UK. Sept. 2015.
- [49] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. *Method for encoding video frames based on local texture synthesis and corresponding device*. France, Patent n° : 15306367.2 - 1908. 2015, European Patent Office (Munich, Germany). 2015.
- [13] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Performance Analysis of Texture Similarity Metrics in HEVC Intra Prediction ». In : *Video Processing and Quality Metrics for Consumer Electronics (VPQM)*. Chandler, Arizona, USA. Fév. 2015.
- [32] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Texture Similarity Metrics Applied to HEVC Intra Prediction ». In : *Sino-French Workshop on Research Collaborations in Informations and Communication Technologies (SIF-WICT)*. Nantes, France. Juin 2015.
- [89] Karam NASER, Vincent RICORDEL et Patrick LE CALLET. « Local texture synthesis : A static texture coding algorithm fully compatible with HEVC ». In : *Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on*. IEEE. 2015, p. 37–40.
- [90] Vincent RICORDEL. *Annual Training Activity Report (ATAR) - Provision ITN*. Livrable D5. Projet ITN PROVISION (FP7-PEOPLE-2013-ITN, agrément n°608231), jan. 2015.
- [5] Vincent RICORDEL, Junle WANG, Matthieu Perreira Da SILVA et Patrick Le CALLET. « 2D & 3D Visual Attention for Computer Vision : Concepts, Measurement, and Modeling ». In : *Innovative Research in Attention Modeling and Computer Vision Applications*. Sous la dir. de Rajarshi PAL. IGI Global publication, 2015. Chap. 1.
- [91] Yi SHEN, Wei DAI et Virginia M RICHARDS. « A MATLAB toolbox for the efficient estimation of the psychometric function using the updated maximum-likelihood adaptive procedure ». In : *Behavior research methods* 47.1 (2015), p. 13–26.

- [92] Fabio SOLARI, Manuela CHESSA, NV Kartheek MEDATHATI et Pierre KORN-PROBST. « What can we expect from a V1-MT feedforward architecture for optical flow estimation? » In : *Signal Processing : Image Communication* 39 (2015), p. 342–354.
- [93] Mathias WIEN. *High Efficiency Video Coding : Coding Tools and Specification*. Signals and Communication Technology. Springer, 2015.
- [94] PP GHADKAR et NB CHOPADE. « Nonlinear Dynamic Texture Analysis and Synthesis Model ». In : *Int. J. of Recent Trends in Engineering & Technology* 11 (2014).
- [95] Christine GUILLEMOT et Olivier LE MEUR. « Image inpainting : Overview and recent advances ». In : *Signal Processing Magazine, IEEE* 31.1 (2014), p. 127–144.
- [96] Hadi HADIZADEH et Ivan V BAJIC. « Saliency-aware video compression ». In : *Image Processing, IEEE Transactions on* 23.1 (2014), p. 19–33.
- [97] Matteo MAGGIONI, Guoxin JIN, Alessandro FOI et Thrasyvoulos N PAPPAS. « Structural texture similarity metric based on intra-class variances ». In : *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2014, p. 1992–1996.
- [98] Rufael N MEKURIA, Pablo CESAR et Dick CA BULTERMAN. « Source coding for transmission of reconstructed dynamic geometry : a rate-distortion-complexity analysis of different approaches ». In : *Applications of Digital Image Processing XXXVII*. T. 9217. International Society for Optics et Photonics, 2014, 92170S.
- [99] Elie Gabriel MORA, Joel JUNG, Marco CAGNAZZO et Béatrice PESQUET-POPESCU. « Depth video coding based on intra mode inheritance from texture ». In : *AP-SIPA Transactions on Signal and Information Processing (Cambridge Univ Press)* 3 (2014).
- [14] Karam NASER, Vincent RICORDEL et Patrick Le CALLET. « Experimenting Texture Similarity Metric STSIM for Intra Prediction Mode Selection and Block Partitioning in HEVC ». In : *Digital Signal Processing (DSP)*. Hong Kong, Chine. Août 2014.
- [100] Ulric NEISSER. *Cognitive Psychology : Classic Edition*. Psychology Press, 2014.
- [101] Carolyn Jeane PERRY et Mazyar FALLAH. « Feature integration and object representations along the dorsal stream visual hierarchy ». In : *Frontiers in computational neuroscience* 8 (2014), p. 84.
- [102] Ruth ROSENHOLTZ. « Texture perception ». In : *The Oxford handbook of perceptual organization* (2014).
- [103] Liuyang WANG, Huaping LIU et Fuchun SUN. « Dynamic texture classification using local fuzzy coding ». In : *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*. IEEE. 2014, p. 1559–1565.

- 
- [104] Miaohui WANG, King Ngi NGAN et Long XU. « Efficient H. 264/AVC Video Coding with Adaptive Transforms ». In : *Multimedia, IEEE Transactions on* 16.4 (2014), p. 933–946.
- [105] Yuanhao ZHAI et David L NEUHOFF. « Rotation-invariant local radius index : A compact texture similarity feature for classification ». In : *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2014, p. 5711–5715.
- [106] Cha ZHANG, Dinei FLORENCIO et Charles LOOP. « Point cloud attribute compression with graph transform ». In : *International Conference on Image Processing (ICIP)*. IEEE. 2014.
- [107] Zhihua BAO, Chen XU et Chong WANG. « Perceptual auto-regressive texture synthesis for video coding ». In : *Multimedia tools and applications* 64.3 (2013), p. 535–547.
- [108] Benjamin BATTIN, Philippe VAUTROT et Marco CAGNAZZO. « Chapitre 10. Codage vidéo multivues ». In : *Vidéo 3D : Capture, traitement et diffusion*. Sous la dir. de Laurent LUCAS, Céline LOSCOS et Yannick REMION. Traité IC2, Série Signal et image. Hermes Lavoisier, sept. 2013.
- [109] Emilie BOSC, Federica BATTISTI, Marco CARLI et Patrick LE CALLET. « A wavelet-based image quality metric for the assessment of 3D synthesized views ». In : *IS&T/SPIE Electronic Imaging*. International Society for Optics et Photonics. 2013, 86481Z.
- [110] Emilie BOSC, Philippe HANHART, Patrick LE CALLET et Touradj EBRAHIMI. « A quality assessment protocol for Free-viewpoint video sequences synthesized from decompressed depth data ». In : *Quality of Multimedia Experience (QoMEX), International Workshop on*. 2013, p. 100–105.
- [111] Emilie BOSC, Patrick LE CALLET, Luce MORIN et Muriel PRESSIGOUT. « Visual quality assessment of synthesized views in the context of 3D-TV ». In : *3D-TV System with Depth-Image-Based Rendering*. Springer, 2013, p. 439–473.
- [112] Emilie BOSC, Fabien RACAPÉ, Vincent JANTET, Paul RIOU, Muriel PRESSIGOUT et Luce MORIN. « A study of depth/texture bit-rate allocation in multi-view video plus depth compression ». In : *Annals of telecommunications-Annales des télécommunications (AT)* 68.11-12 (2013), p. 615–625.
- [113] John Horton CONWAY et Neil James Alexander SLOANE. *Sphere packings, lattices and groups*. T. 290. Springer Science & Business Media, 2013.
- [114] Tomás CRIVELLI, Bruno CERNUSCHI-FRIAS, Patrick BOUTHEMY et Jian-Feng YAO. « Motion textures : Modeling, classification, and segmentation using mixed-state Markov random fields ». In : *SIAM Journal on Imaging Sciences* 6.4 (2013), p. 2484–2520.

- 
- [115] Konstantinos G DERPANIS, Mikhail SIZINTSEV, Kevin J CANNONS et Richard P WILDES. « Action spotting and recognition based on a spatiotemporal orientation analysis ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.3 (2013), p. 527–540.
- [116] Dan GROIS, Detlev MARPE, Amit MULAYOFF, Benaya ITZHAKY et Ofer HADAR. « Performance comparison of H. 265/MPEG-HEVC, VP9, and H. 264/MPEG-AVC encoders ». In : *Picture Coding Symposium (PCS)*. IEEE. 2013, p. 394–397.
- [117] C. GUILLEMOT et L. GUILLO. *Additional merge candidates derived from shifted disparity candidate predictors*. Joint Collaborative Team On 3D Video Coding Extension Development, JCT3V-C148. 2013.
- [118] Christine GUILLEMOT, Mehmet TURKAN, Olivier LE MEUR et Mounira EBDELLI. « Image inpainting using LLE-LDNR and linear subspace mappings ». In : *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*. IEEE. 2013, p. 1558–1562.
- [119] Yimo GUO, Guoying ZHAO, Ziheng ZHOU et Matti PIETIKAINEN. « Video texture synthesis with multi-frame LBP-TOP and diffeomorphic growth model ». In : *Image Processing, IEEE Transactions on* 22.10 (2013), p. 3879–3891.
- [120] Hadi HADIZADEH. « Visual saliency in video compression and transmission ». Thèse de doct. Applied Sciences : School of Engineering Science, 2013.
- [121] Michael S LANDY. « Texture analysis and perception ». In : *The new visual neurosciences, MIT Press, Cambridge, Mass* (2013), p. 639–652.
- [122] Siwei MA, Shiqi WANG, Qin YU, Junjun Si et Wen GAO. « Mode dependent coding tools for video coding ». In : *Selected Topics in Signal Processing, IEEE Journal of* 7.6 (2013), p. 990–1000.
- [123] E. G. MORA, J JUNG, Marco CAGNAZZO et Béatrice PESQUET-POPESCU. « Modification of the merge candidate list for dependent views in 3D-HEVC ». In : *Image Processing (ICIP), IEEE International Conference on*. IEEE. 2013, p. 1709–1713.
- [124] Elie Gabriel MORA, Joel JUNG, Beatrice PESQUET-POPESCU et Marco CAGNAZZO. « Modification of the disparity vector derivation process in 3D-HEVC ». In : *Multimedia Signal Processing (MMSP), IEEE International Workshop on*. IEEE. 2013, p. 206–211.
- [125] Elie Gabriel MORA, Joël JUNG, Béatrice PESQUET-POPESCU et Marco CAGNAZZO. « Chapitre 12. Méthodes de codage de vidéos de profondeur ». In : *Vidéo 3D : Capture, traitement et diffusion*. Sous la dir. de Laurent LUCAS, Céline LOSCOS et Yannick REMION. Traité IC2, Série Signal et image. Hermes Lavoisier, sept. 2013.



- [126] Luce MORIN, Olivier LE MEUR, Christine GUILLEMOT et Vincent JAN-TET AUTIER. « Chapitre 9. Synthèse de vue intermédiaire ». In : *Vidéo 3D : Capture, traitement et diffusion*. Sous la dir. de Laurent LUCAS, Céline LOS-COS et Yannick REMION. Traité IC2, Série Signal et image. Hermes Lavoisier, sept. 2013.
- [127] J-R. OHM et Gary J. SULLIVAN. « High efficiency video coding : the next frontier in video compression [Standards in a Nutshell]. » In : *Signal Processing Magazine, IEEE* 30.1 (2013), p. 152–158.
- [128] Liquan SHEN, Zhi LIU et Zhaoyang ZHANG. « A novel H.264 rate control algorithm with consideration of visual attention ». In : *Multimedia tools and applications* 63.3 (2013), p. 709–727.
- [129] Mehmet TÜRKAN et Christine GUILLEMOT. « Dictionary learning for image prediction ». In : *Journal of Visual Communication and Image Representation (Elsevier)* 24.3 (2013), p. 426–437.
- [130] Junle WANG, Josselin GAUTIER, Olivier Le MEUR, Emilie BOSC et Vincent RICORDEL. *Perceptual Modelling : Softwares results and final report*. Livrable D1.3. Projet PERSEE (ANR-09-BLAN-0170), juil. 2013.
- [2] Junle WANG, Matthieu Perreira Da SILVA, Patrick Le CALLET et Vincent RICORDEL. « A computational model of stereoscopic 3D visual saliency ». In : *IEEE Transaction on Image Processing* 22.6 (juin 2013), p. 2151–2165.
- [15] Junle WANG, Matthieu Perreira Da SILVA, Patrick Le CALLET et Vincent RICORDEL. « Study of center-bias in the viewing of stereoscopic image and a framework for extending 2D visual attention models to 3D ». In : *SPIE Electronic Imaging*. San Francisco, USA. Fév. 2013, p. 3–7.
- [131] Andrew B WATSON. « High frame rates and human vision : A view through the window of visibility ». In : *SMPTE Motion Imaging Journal* 122.2 (2013), p. 18–32.
- [132] David WOLINSKI, Olivier LE MEUR et Josselin GAUTIER. « 3D view synthesis with inter-view consistency ». In : *Proceedings of the 21st ACM international conference on Multimedia*. ACM. 2013, p. 669–672.
- [133] Yuanhao ZHAI, David L NEUHOFF et Thrasyvoulos N PAPPAS. « Local radius index-a new texture similarity feature ». In : *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE. 2013, p. 1434–1438.
- [134] Jana ZUJOVIC, Thrasyvoulos N PAPPAS et David L NEUHOFF. « Structural texture similarity metrics for image analysis and retrieval ». In : *Image Processing, IEEE Transactions on* 22.7 (2013), p. 2545–2558.
- [135] John G APOSTOLOPOULOS, Philip A CHOU, Bruce CULBERTSON, Ton KALKER, Mitchell D TROTT et Susie WEE. « The road to immersive communication ». In : *Proceedings of the IEEE*. T. 100. 4. IEEE. 2012, p. 974–990.

- [136] Emilie BOSC. « Compression of Multi-View-plus-Depth (MVD) data : from perceived quality analysis to MVD coding tools designing ». Thèse de doct. INSA de Rennes, oct. 2012.
- [137] Emilie BOSC, Patrick LE CALLET, Luce MORIN et Muriel PRESSIGOUT. « An edge-based structural distortion indicator for the quality assessment of 3D synthesized views ». In : *Picture Coding Symposium (PCS)*. IEEE. 2012, p. 249–252.
- [138] Emilie BOSC, Luce MORIN et Muriel PRESSIGOUT. « A content-based method for perceptually driven joint color/depth compression ». In : *IS&T/SPIE Electronic Imaging*. International Society for Optics et Photonics. 2012, p. 82882C.
- [139] Emilie BOSC, Romuald PÉPION, Patrick LE CALLET, Muriel PRESSIGOUT et Luce MORIN. « Reliability of 2D quality assessment methods for synthesized views evaluation in stereoscopic viewing conditions ». In : *3DTV-Conference : The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*. IEEE. 2012, p. 1–4.
- [140] Emilie BOSC, Paul RIOU, Muriel PRESSIGOUT et Luce MORIN. « Bit-rate allocation between texture and depth : influence of data sequence characteristics ». In : *3DTV-Conference : The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*. IEEE. 2012, p. 1–4.
- [141] Konstantinos G DERPANIS et Richard P WILDES. « Spacetime texture representation and recognition based on a spatiotemporal orientation analysis ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34.6 (2012), p. 1193–1205.
- [142] Jie DONG et King Ngi NGAN. « Two-layer directional transform for high performance video coding ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 22.4 (2012), p. 619–625.
- [143] Peter FLACH. *Machine learning : the art and science of algorithms that make sense of data*. Cambridge University Press, 2012.
- [144] J GAUTIER et O LE MEUR. « Un modèle de saillance dépendant du temps combinant les biais centré et de profondeur pour la visualisation en 2D et 3D ». In : *CORESA*. 2012.
- [145] Josselin GAUTIER. « A Dynamic Visual Attention Model for 2D and 3D conditions ; Depth Coding and Inpainting-based Synthesis for Multiview Videos ». Thèse de doct. Université Rennes 1, déc. 2012.
- [146] Josselin GAUTIER et Olivier LE MEUR. « A time-dependent saliency model combining center and depth biases for 2D and 3D viewing conditions ». In : *Cognitive Computation (Springer)* 4.2 (2012), p. 141–156.
- [147] Josselin GAUTIER, Olivier LE MEUR et Christine GUILLEMOT. « Efficient depth map compression based on lossless edge coding and diffusion ». In : *Picture Coding Symposium (PCS)*. IEEE. 2012, p. 81–84.

- [148] Wesley Nunes GONÇALVES, Bruno Brandoli MACHADO et Odemir Martinez BRUNO. « Spatiotemporal Gabor filters : a new method for dynamic texture recognition ». In : *arXiv preprint arXiv :1201.3612* (2012).
- [149] T. GUIONNET, L. GUILLO et C. GUILLEMOT. *Merge candidate list for disparity compensated prediction*. Joint Collaborative Team On 3D Video Coding Extension Development, JCT3V-B0080. 2012.
- [150] Multi-Partner ITN. *PeRceptually Optimised Video CompresSION (PROVISION) - part A*. Proposal to : Marie Curie Initial Training Networks (call FP7-PEOPLE-2013-ITN). 2012.
- [151] Multi-Partner ITN. *PeRceptually Optimised Video CompresSION (PROVISION) - part B*. Proposal to : Marie Curie Initial Training Networks (call FP7-PEOPLE-2013-ITN). 2012.
- [152] Vincent JANTET. « Layered Depth Images for Multi-View Coding ». Thèse de doct. Université Rennes 1, nov. 2012.
- [153] Julius KAMMERL, Nico BLOWOW, Radu Bogdan RUSU, Suat GEDIKLI, Michael BEETZ et Eckehard STEINBACH. « Real-time compression of point cloud streams ». In : *International Conference on Robotics and Automation*. IEEE. 2012, p. 778–785.
- [154] Olivier LE MEUR et Christine GUILLEMOT. « Super-resolution-based inpainting ». In : *Computer Vision—ECCV 2012*. Springer, 2012, p. 554–567.
- [155] Jing LI, Marcus BARKOWSKY et Patrick LE CALLET. « Analysis and improvement of a paired comparison method in the application of 3DTV subjective experiment ». In : *Image Processing (ICIP), IEEE International Conference on*. IEEE. 2012, p. 629–632.
- [156] Meifeng LIU et Ling LU. « An Improved Rate Control Algorithm of H.264/AVC Based on Human Visual System ». In : *Computer, Informatics, Cybernetics and Applications*. Springer, 2012, p. 1145–1151.
- [157] E. MORA, B. PESQUET, M. CAGNAZZO et J. JUNG. *Modification of the Merge Candidate List for Dependant Views in 3DV-HTM*. JCT3V-B0069 for Shanghai meeting (MPEG number M26793), Shangai (PRC). 2012.
- [158] EG MORA, J JUNG, M CAGNAZZO et B PESQUET-POPESCU. « Codage de vidéos de profondeur basé sur l’héritage des modes intra de texture ». In : *CORESA*. T. 1. 2012, p. 1–4.
- [159] Patrick NDJIKI-NYA, Dimitar DOSHKOV, Hagen KAPRYKOWSKY, Fang ZHANG, Dave BULL et Thomas WIEGAND. « Perception-oriented video coding based on image analysis and completion : A review ». In : *Signal Processing : Image Communication (Elsevier)* 27.6 (2012), p. 579–594.

- 
- [160] J. R. OHM, G. J. SULLIVAN, H. SCHWARZ, T. K. TAN et T. WIEGAND. « Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC) ». In : *IEEE Transactions on Circuits and Systems for Video Technology* 22.12 (2012), p. 1669–1684.
- [161] Jens-Rainer OHM. *Multimedia communication technology : Representation, transmission and identification of multimedia signals*. Springer Science & Business Media, 2012.
- [162] Yi SHEN et Virginia M RICHARDS. « A maximum-likelihood procedure for estimating psychometric functions : Thresholds, slopes, and lapses of attention ». In : *The Journal of the Acoustical Society of America* 132.2 (2012), p. 957–967.
- [163] David I SHUMAN, Sunil K NARANG, Pascal FROSSARD, Antonio ORTEGA et Pierre VANDERGHEYNST. « The emerging field of signal processing on graphs : Extending high-dimensional data analysis to networks and other irregular domains ». In : *arXiv preprint arXiv :1211.0053* (2012).
- [164] G. J. SULLIVAN, J. R. OHM, W. J. HAN et T. WIEGAND. « Overview of the High Efficiency Video Coding (HEVC) Standard ». In : *IEEE Transactions on Circuits and Systems for Video Technology* 22.12 (2012), p. 1649–1668.
- [165] Gary J SULLIVAN, Jens OHM, Woo-Jin HAN et Thomas WIEGAND. « Overview of the high efficiency video coding (HEVC) standard ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 22.12 (2012), p. 1649–1668.
- [166] David TAUBMAN et Michael MARCELLIN. *JPEG2000 Image Compression Fundamentals, Standards and Practice : Image Compression Fundamentals, Standards and Practice*. Springer Science & Business Media, 2012.
- [167] V. THIRUMALAI, L. ZHANG, Y. CHEN, M. KARCZEWICZ, T. GUIONNET, C. GUILLEMOT et L. GUILLO. *Merge candidates derivation from vector shifting*. Joint Collaborative Team On 3D Video Coding Extension Development, JCT3V-D0178. 2012.
- [168] Mehmet TURKAN et Christine GUILLEMOT. « Locally linear embedding based texture synthesis for image prediction and error concealment ». In : *Image Processing (ICIP), IEEE International Conference on*. IEEE. 2012, p. 3009–3012.
- [169] Mehmet TÜRKAN et Christine GUILLEMOT. « Image prediction based on neighbor-embedding methods ». In : *Image Processing, IEEE Transactions on* 21.4 (2012), p. 1885–1898.
- [16] Matthieu URVOY, Marcus BARKOWSKY, Romain COUSSEAU, Yao KOUDOTA, Vincent RICORDEL, Patrick Le Calletand Jesús GUTIÉRREZ et Narciso GARCIA. « NAMA3DS1-COSPAD1 : Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences ». In : *International Workshop on Quality of Multimedia Experience (QoMEX)*. Yarra Valley, Australie. Juil. 2012.

- [170] Giuseppe VALENZISE, Gene CHEUNG, R GALVAO, Marco CAGNAZZO, Béatrice PESQUET-POPESCU et Antonio ORTEGA. « Motion prediction of depth video for depth-image-based rendering using don't care regions ». In : *Picture Coding Symposium (PCS)*. IEEE. 2012, p. 93–96.
- [171] Junle WANG. « From 2D to Stereoscopic-3D Visual Saliency : Revisiting Psychophysical Methods and Computational Modeling ». Thèse de doct. L'Université Nantes Angers Le Mans (LUNAM), nov. 2012.
- [3] Junle WANG, Patrick Le CALLET, Sylvain TOURANCHEAU, Vincent RICORDEL et Matthieu Pereira Da SILVA. « Study of depth bias of observers in free viewing of still stereoscopic synthetic stimuli ». In : *Journal of Eye Movement Research* 5.5 (sept. 2012), p. 1–11.
- [172] Shiqi WANG, Abdul REHMAN, Zhou WANG, Siwei MA et Wen GAO. « SSIM-motivated rate-distortion optimization for video coding ». In : *IEEE Transactions on Circuits and Systems for Video Technology* 22.4 (2012), p. 516–529.
- [173] Long XU, King Ngi NGAN et Miaohui WANG. « Video content dependent directional transform for intra frame coding ». In : *Picture Coding Symposium (PCS), 2012*. IEEE. 2012, p. 197–200.
- [174] Chuohao YEO, Yih Han TAN, Zhengguo LI et Susanto RAHARDJA. « Mode-dependent transforms for coding directional intra prediction residuals ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 22.4 (2012), p. 545–554.
- [175] Elena ALSHINA, Alexander ALSHIN et Felix C FERNANDES. « Rotational transform for image and video compression ». In : *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE. 2011, p. 3689–3692.
- [176] Emilie BOSC, Vincent JANTET, Muriel PRESSIGOUT, Luce MORIN et Christine GUILLEMOT. « Bit-rate allocation for multi-view video plus depth ». In : *3DTV Conference : The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*. Sous la dir. de 1–4. IEEE. 2011.
- [177] Emilie BOSC, Martin KOPPEL, Romuald PÉPION, Muriel PRESSIGOUT, Luce MORIN, Patrick NDJIKI-NYA et Patrick LE CALLET. « Can 3D synthesized views be reliably assessed through usual subjective and objective evaluation protocols? » In : *Image Processing (ICIP), IEEE International Conference on*. IEEE. 2011, p. 2597–2600.
- [178] Emilie BOSC, Romuald PÉPION, Patrick LE CALLET, Martin KÖPPEL, Patrick NDJIKI-NYA, Luce MORIN et Muriel PRESSIGOUT. « Perceived quality of DIBR-based synthesized views ». In : *SPIE Optical Engineering+ Application*. International Society for Optics et Photonics. 2011, p. 81350I.
- [179] Emilie BOSC, Romuald PÉPION, Patrick LE CALLET, Martin KÖPPEL, Patrick NDJIKI-NYA, Muriel PRESSIGOUT et Luce MORIN. « Towards a new quality metric for 3-D synthesized view assessment ». In : *Selected Topics in Signal Processing, IEEE Journal of*. T. 5. 7. IEEE, 2011, p. 1332–1343.

- 
- [180] Emilie BOSC, Romuald PEPION, Patrick LE CALLET, Martin KÖPPEL, Patrick NDJIKI-NYA, Muriel PRESSIGOUT et Luce MORIN. « Towards a new quality metric for 3-D synthesized view assessment ». In : *Selected Topics in Signal Processing, IEEE Journal of* 5.7 (2011), p. 1332–1343.
- [181] Emilie BOSC, Muriel PRESSIGOUT et Luce MORIN. « 3D Video : new techniques and challenges ». In : *First Sino-French Workshop on Education and Research collaborations in Information and Communication Technologies (SIFWICT)*. 2011.
- [182] Emilie BOSC, Muriel PRESSIGOUT et Luce MORIN. « Évaluation de la qualité des vues 3D synthétisées ». In : *ORASIS*. Praz-sur-Arly, France. 2011, p. 39.
- [183] Marco CAGNAZZO et Béatrice PESQUET-POPESCU. « Depth map coding by dense disparity estimation for mvd compression ». In : *Digital Signal Processing (DSP), International Conference on*. IEEE. 2011, p. 1–6.
- [184] Rafael Galvão DE OLIVEIRA, Béatrice PESQUET-POPESCU et Maria TROCAN. « Inter prediction using lapped transforms for advanced video coding ». In : *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE. 2011, p. 3705–3708.
- [185] Josselin GAUTIER, Olivier LE MEUR et Christine GUILLEMOT. « Depth-based image completion for view synthesis ». In : *3DTV Conference : The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*. T. 1–4. IEEE. 2011.
- [186] T. GUIONNET et L. GUILLO. *Intra prediction based on weighed template matching predictors (WTM)*. Joint Collaborative Team On Video Coding, JCTVC-G598. 2011.
- [187] Quan HUYNH-THU, Marcus BARKOWSKY et Patrick Le CALLET. « The importance of visual attention in improving the 3D-TV viewing experience : Overview and new perspectives ». In : *Broadcasting, IEEE Transactions on* 57.2 (2011), p. 421–431.
- [188] Vincent JANTET, Christine GUILLEMOT et Luce MORIN. « Joint projection filling method for occlusion handling in depth-image-based rendering ». In : *3D Research* 2.4 (2011), p. 1–13.
- [189] Vincent JANTET, Christine GUILLEMOT et Luce MORIN. « Object-based Layered Depth Images for improved virtual view synthesis in rate-constrained context ». In : *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE. 2011, p. 125–128.
- [190] R JOSHI, P CHEN, M KARCZEWICZ, A TANIZAWA, J YAMAGUCHI, C YEO, YH TAN, H YANG et H YU. « CE7 : Mode dependent intra residual coding ». In : *document JCTVC-E098, MPEG-H/JCT-VC, Geneva, Switzerland* (2011).

- 
- [191] Hagen KAPRYKOWSKY, Dimitar DOSHKOV, Christoph HOFFMANN, Patrick NDJIKI-NYA et Thomas WIEGAND. « Investigation of perception-oriented coding techniques for video compression based on large block structures ». In : *SPIE Optical Engineering+ Applications*. International Society for Optics et Photonics. 2011, 81350K.
- [192] Olivier LE MEUR. « Predicting saliency using two contextual priors : the dominant depth and the horizon line ». In : *Multimedia and Expo (ICME), IEEE International Conference on*. IEEE. 2011, p. 1–6.
- [193] Olivier LE MEUR, Josselin GAUTIER et Christine GUILLEMOT. « Exemplar-based inpainting based on local geometry ». In : *Image Processing (ICIP), IEEE International Conference on*. IEEE. 2011, p. 3401–3404.
- [194] Jing LI, Marcus BARKOWSKY et Patrick LE CALLET. « The influence of relative disparity and planar motion velocity on visual discomfort of stereoscopic videos ». In : *Quality of multimedia experience (QoMEX), International workshop on*. IEEE. 2011, p. 155–160.
- [195] Jing LI, Marcus BARKOWSKY, Junle WANG et Patrick LE CALLET. « Study on visual discomfort induced by stimulus movement at fixed depth on stereoscopic displays using shutter glasses ». In : *Digital Signal Processing (DSP), International Conference on*. IEEE. 2011, p. 1–8.
- [196] Hantao LIU. « Modeling Perceived Quality for Imaging Applications ». Thèse de doct. TU Delft, Delft University of Technology, 2011.
- [197] Hantao LIU, Junle WANG, Judith REDI, Patrick Le CALLET et Ingrid HEYNDE-RICKX. « An efficient no-reference metric for perceived blur ». In : *Visual Information Processing (EUVIP), European Workshop on*. IEEE. 2011, p. 174–179.
- [198] Shinji NISHIMOTO et Jack L GALLANT. « A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies ». In : *The Journal of Neuroscience* 31.41 (2011), p. 14551–14564.
- [199] Ekaterina POTAPOVA, Michael ZILLICH et Markus VINCZE. « Learning what matters : combining probabilistic models of 2D and 3D saliency cues ». In : *Computer Vision Systems*. Springer, 2011, p. 132–142.
- [33] Cédric RAMASSAMY, Vincent RICORDEL, Oana CRAMARIUC et Bogdan CRAMARIUC. « Lattice Vector Quantization for the Analysis of Molecular Data ». In : *Signal Processing and Applied Mathematics for Electronics and Communications (SPAMEC)*. Cluj-Napoca, Roumanie. Août 2011.
- [200] Radu B RUSU et S COUSINS. « Point cloud library (pcl) ». In : *2011 IEEE International Conference on Robotics and Automation*. 2011, p. 1–4.
- [201] Ankur SAXENA et Felix C FERNANDES. « Mode dependent DCT/DST for intra prediction in block-based image/video coding ». In : *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE. 2011, p. 1685–1688.

- [202] Osman G SEZER, Robert COHEN et Anthony VETRO. « Robust learning of 2-D separable transforms for next-generation video coding ». In : *Data Compression Conference (DCC)*. IEEE. 2011, p. 63–72.
- [203] Emilien TLAPALE, Pierre KORNPORST, Guillaume S MASSON et Olivier FAUGERAS. « A neural field model for motion estimation ». In : *Mathematical image processing*. Springer, 2011, p. 159–179.
- [204] Mehmet TURKAN et Christine GUILLEMOT. « Online dictionaries for image prediction ». In : *Image Processing (ICIP), IEEE International Conference on*. IEEE. 2011, p. 293–296.
- [205] Dong WANG, Guiqing LI, Weijia JIA et Xiaonan LUO. « Saliency-driven scaling optimization for image retargeting ». In : *The Visual Computer*. T. 27. 9. Springer, 2011, p. 853–860.
- [34] Junle WANG, Marcus BARKOWSKY, Vincent RICORDEL et Patrick Le CALLET. « Quantifying how the combination of blur and disparity affects the perceived depth ». In : *Sino-French Workshop on Research Collaborations in Informations and Communication Technologies (SIFWICT)*. Nantes, France. Mai 2011.
- [17] Junle WANG, Marcus BARKOWSKY, Vincent RICORDEL et Patrick Le CALLET. « Quantifying how the combination of blur and disparity affects the perceived depth ». In : *SPIE Electronic Imaging*. San Francisco Airport, California, USA. Jan. 2011.
- [206] Junle WANG, Emilie BOSC, Jing LI et Vincent RICORDEL. *Perceptual Modelling : Definition of the Models*. Livrable D1.2. Projet PERSEE (ANR-09-BLAN-0170), oct. 2011.
- [18] Junle WANG, Patrick Le CALLET, Vincent RICORDEL et Sylvain TOURANCHEAU. « Quantifying depth bias in free viewing of still stereoscopic synthetic stimuli ». In : *European Conference on Eye Movements (ECEM)*. Marseille, France. Août 2011.
- [207] Junle WANG, Josselin GAUTIER, Emilie BOSC, Jing LI et Vincent RICORDEL. *Perceptual Assessment : Definition of the scenarios*. Livrable D6.1. Projet PERSEE (ANR-09-BLAN-0170), oct. 2011.
- [208] Thomas WIEGAND, Woo-Jin HAN, Benjamin BROSS, Jens-Rainer OHM et Gary J SULLIVAN. « WD3 : Working draft 3 of high-efficiency video coding ». In : *JCT-VC 5th Meeting, JCTVC-E603*. 2011.
- [209] Chuohao YEO, Yih Han TAN, Zhengguo LI et Susanto RAHARDJA. « Mode-dependent fast separable KLT for block-based intra coding ». In : *Circuits and Systems (ISCAS), 2011 IEEE International Symposium on*. IEEE. 2011, p. 621–624.
- [210] Fan ZHANG et David R BULL. « A parametric framework for video compression using region-based texture models ». In : *Selected Topics in Signal Processing, IEEE Journal of* 5.7 (2011), p. 1378–1392.



- [211] Fan ZHANG et David R BULL. « A parametric framework for video compression using region-based texture models ». In : *Selected Topics in Signal Processing, IEEE Journal of* 5.7 (2011), p. 1378–1392.
- [212] Qi ZHAO et Christof KOCH. « Learning a saliency map using fixated locations in natural scenes ». In : *Journal of vision* 11.3 (2011), p. 9.
- [213] Feng ZOU, Oscar C AU, Chao PANG et Jingjing DAI. « Rate distortion optimized transform for intra block coding for HEVC ». In : *Visual Communications and Image Processing (VCIP), 2011 IEEE*. IEEE. 2011, p. 1–4.
- [214] Moyuresh BISWAS, Mark R PICKERING et Michael R FRATER. « Improved H.264-based video coding using an adaptive transform ». In : *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE. 2010, p. 165–168.
- [215] Emilie BOSC, Vincent JANTET, Luce MORIN, Muriel PRESSIGOUT et Christine GUILLEMOT. « Vidéo 3D : quel débit pour la profondeur ? » In : *CORESA*. Lyon, France. 2010.
- [216] Emilie BOSC, Muriel PRESSIGOUT et Luce MORIN. « Focus on visual rendering quality through content-based depth map coding ». In : *Picture Coding Symposium (PCS)*. IEEE. 2010, p. 158–161.
- [217] Olivier BROUARD. « Pré-analyse de la vidéo pour un codage adapté. Application au codage de la TVHD en flux H.264 ». Thèse de doct. Université de Nantes, 2010.
- [218] Madhukar BUDAGAVI et Minhua ZHOU. « Orthogonal MDDT and mode dependent DCT ». In : *ITU-T Q 6* (2010).
- [219] M CAGNAZZO et B PESQUET-POPESCU. « Introducing differential motion estimation into hybrid video coders ». In : *Visual Communications and Image Processing (VCIP)*. International Society for Optics et Photonics. 2010, p. 774415.
- [220] Marco CAGNAZZO et Béatrice PESQUET-POPESCU. « Perceptual impact of transform coefficients quantization for adaptive lifting schemes ». In : *International Workshop on Video Processing and Quality Metrics for Consumer Electronics*. T. 160. 2010.
- [221] Christel CHAMARET, Sylvain GODEFFROY, Patrick LOPEZ et Olivier LE MEUR. « Adaptive 3D rendering based on region-of-interest ». In : *IS&T/SPIE Electronic Imaging*. ISOP. 2010, p. 75240V.
- [222] Chuo-Ling CHANG, Mina MAKAR, Sam S TSAI et Bernd GIROD. « Direction-adaptive partitioned block transform for color image coding ». In : *Image Processing, IEEE Transactions on* 19.7 (2010), p. 1740–1755.
- [223] Homer H CHEN, Yi-Hsin HUANG, Po-Yen SU et Tao-Sheng OU. « Improving video coding quality by perceptual rate-distortion optimization ». In : *Multimedia and Expo (ICME), 2010 IEEE International Conference on*. IEEE. 2010, p. 1287–1292.

- 
- [224] Jie CHEN, Shiguang SHAN, Chu HE, Guoying ZHAO, Matti PIETIKAINEN, Xilin CHEN et Wen GAO. « WLD : A robust local image descriptor ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32.9 (2010), p. 1705–1720.
- [225] Robert COHEN, Sven KLOMP, Anthony VETRO, Huifang SUN et al. « Direction-adaptive transforms for coding prediction residuals ». In : *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE. 2010, p. 185–188.
- [226] Konstantinos G DERPANIS et Richard P WILDES. « Dynamic texture recognition based on distributions of spacetime oriented structure ». In : *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, p. 191–198.
- [227] Angelique DREMEAU, Mehmet TURKAN, Cedric HERZET, Christine GUILLEMOT et Jean-Jacques FUCHS. « Spatial intra-prediction based on mixtures of sparse representations ». In : *Multimedia Signal Processing (MMSP), IEEE International Workshop on*. IEEE. 2010, p. 345–349.
- [228] Ulrich ENGELKE, Anthony MAEDER et Hans-Jürgen ZEPERNICK. « Analysing inter-observer saliency variations in task-free viewing of natural images ». In : *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE. 2010, p. 1085–1088.
- [229] Bernard GHANEM et Narendra AHUJA. « Maximum margin distance learning for dynamic texture recognition ». In : *European Conference on Computer Vision*. Springer. 2010, p. 223–236.
- [230] Milind S GIDE et Lina J KARAM. « On the assessment of the quality of textures in visual media ». In : *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*. IEEE. 2010, p. 1–5.
- [231] Christine GUILLEMOT, Laurent GUILLO et Tangi POIRIER. *Intra Prediction by a linear combination of Template Matching predictors*. Joint Collaborative Team on Video Coding, JCTVC-B078. 2010.
- [232] Jukka HÄKKINEN, Takashi KAWAI, Jari TAKATALO, Reiko MITSUYA et Göte NYMAN. « What do people look at when they watch stereoscopic movies? » In : *IS&T/SPIE Electronic Imaging*. International Society for Optics et Photonics. 2010, 75240E.
- [233] Dong-Chen HE et Li WANG. « Simplified texture spectrum for texture analysis ». In : *Journal of Communication and Computer* 7.8 (2010), p. 44–53.
- [234] Robert T HELD, Emily A COOPER, James F O’BRIEN et Martin S BANKS. « Using blur to affect perceived distance and size ». In : *ACM transactions on graphics* 29.2 (2010).
- [235] David M HOFFMAN et Martin S BANKS. « Focus information is used to interpret binocular images ». In : *Journal of vision* 10.5 (2010), p. 13.

- [236] Yi-Hsin HUANG, Tao-Sheng OU, Po-Yen SU et Homer H CHEN. « Perceptual rate-distortion optimization using structural similarity index as quality metric ». In : *IEEE Transactions on Circuits and Systems for Video Technology* 20.11 (2010), p. 1614–1624.
- [237] Sophie MARAT. « Modèles de saillance visuelle par fusion d’informations sur la luminance, le mouvement et les visages pour la prédiction de mouvements oculaires lors de l’exploration de vidéos ». Thèse de doct. Université Joseph-Fourier-Grenoble I, 2010.
- [238] Philipp MERKLE, Karsten MÜLLER et Thomas WIEGAND. « 3D video coding : an overview of present and upcoming standards ». In : *Visual Communications and Image Processing (VCIP)*. International Society for Optics et Photonics. 2010, p. 77440D–77440D.
- [239] Renaud PÉTERI, Sándor FAZEKAS et Mark J HUISKES. « DynTex : A comprehensive database of dynamic textures ». In : *Pattern Recognition Letters* 31.12 (2010), p. 1627–1632.
- [240] Vincent RICORDEL, Junle WANG, Josselin GAUTIER, Olivier Le MEUR et Emilie BOSC. *Perceptual modelling for 2D and 3D*. Livrable D1.1. Projet PERSEE (ANR-09-BLAN-0170), nov. 2010.
- [241] Kalpana SESHADRINATHAN et Alan Conrad BOVIK. « Motion tuned spatio-temporal quality assessment of natural videos ». In : *Image Processing, IEEE Transactions on* 19.2 (2010), p. 335–350.
- [242] Junle WANG, Damon M CHANDLER et Patrick LE CALLET. « Quantifying the relationship between visual salience and visual importance ». In : *IS&T/SPIE Electronic Imaging*. International Society for Optics et Photonics. 2010, 75270K.
- [243] Yun ZHANG, Gangyi JIANG, Mei YU et Ken CHEN. « Stereoscopic visual attention model for 3d video ». In : *Advances in Multimedia Modeling*. Springer, 2010, p. 314–324.
- [244] Xin ZHAO, Li ZHANG, Siwei MA et Wen GAO. « Rate-distortion optimized transform for intra-frame coding ». In : *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE. 2010, p. 1414–1417.
- [245] Ravi ACHANTA, Sheila HEMAMI, Francisco ESTRADA et Sabine SUSSTRUNK. « Frequency-tuned salient region detection ». In : *Computer vision and pattern recognition, 2009. cvpr 2009. iee conference on*. IEEE. 2009, p. 1597–1604.
- [246] Marc BOSCH, Fengqing ZHU et Edward J DELP. « An overview of texture and motion based video coding at Purdue University ». In : *Picture Coding Symposium, 2009. PCS 2009*. IEEE. 2009, p. 1–4.
- [38] Olivier BROUARD, Vincent RICORDEL et Dominique BARBA. « Cartes de Saillance Spatio-Temporelle basées Contrastes de Couleur et Mouvement Relatif ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Toulouse, France. Mar. 2009.

- [247] Neil DB BRUCE et John K TSOTSOS. « Saliency, attention, and visual search : An information theoretic approach ». In : *Journal of vision* 9.3 (2009), p. 5.
- [248] Sloven DUBOIS, Renaud PÉTERI et Michel MÉNARD. « A comparison of wavelet based spatio-temporal decomposition methods for dynamic texture recognition ». In : *Pattern Recognition and Image Analysis*. Springer, 2009, p. 314–321.
- [249] Ulrich ENGELKE, Anthony MAEDER et Hans-Jürgen ZEPERNICK. « Visual attention modelling for subjective image quality databases ». In : *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*. IEEE. 2009, p. 1–6.
- [250] Sándor FAZEKAS, Tomer AMIAZ, Dmitry CHETVERIKOV et Nahum KIRYATI. « Dynamic texture detection based on motion analysis ». In : *International journal of computer vision* 82.1 (2009), p. 48–63.
- [251] Stephen IERODIACONOU, James BYRNE, David R BULL, David REDMILL et Paul HILL. « Unsupervised image compression using graphcut texture synthesis ». In : *Proceedings of the 16th IEEE international conference on Image processing*. IEEE. 2009, p. 2265–2268.
- [252] Lina JANSEN, Selim ONAT et Peter KÖNIG. « Influence of disparity on fixation and saccades in free viewing of natural scenes ». In : *Journal of Vision* 9.1 (2009), p. 29.
- [39] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « Étude comparative de techniques de lifting orientées sur des contenus synthétiques représentant des structures rectilignes ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Toulouse, France. Mar. 2009.
- [253] Helga KOLB, Eduardo FERNANDEZ, Ralph NELSON et Helga KOLB. « Neurotransmitters in the Retina ». In : *University of Utah Health Sciences Center* (2009).
- [254] Hantao LIU et Ingrid HEYNDERICKX. « Studying the added value of visual attention in objective image quality metrics based on eye movement data ». In : *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE. 2009, p. 3097–3100.
- [255] Khaled MAMOU, Titus ZAHARIA et Françoise PRÊTEUX. « TFAN : A low complexity 3D mesh compression algorithm ». In : *Computer Animation and Virtual Worlds* 20.2-3 (2009), p. 343–354.
- [256] Sophie MARAT, Nathalie GUYADER et Denis PELLERIN. « Gaze prediction improvement by adding a face feature to a saliency model ». In : *Recent advances in signal processing* (2009), p. 195–210.
- [257] Sophie MARAT, Tien Ho PHUOC, Lionel GRANJON, Nathalie GUYADER, Denis PELLERIN et Anne GUÉRIN-DUGUÉ. « Modelling spatio-temporal saliency to predict gaze direction for short videos ». In : *International journal of computer vision* 82.3 (2009), p. 231–243.

- 
- [258] George MATHER. *Foundations of Sensation and Perception*. Psychology Press, 2009.
- [259] Patrick NDJIKI-NYA, Dave BULL et Thomas WIEGAND. « Perception-oriented video coding based on texture analysis and synthesis ». In : *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE. 2009, p. 2273–2276.
- [260] Celambarasan RAMASAMY, Donald H HOUSE, Andrew T DUCHOWSKI et Brian DAUGHERTY. « Using eye tracking to analyze stereoscopic filmmaking ». In : *SIGGRAPH'09 : Posters*. ACM. 2009, p. 28.
- [261] Joel SOLE, Peng YIN, Yunfei ZHENG et Cristina GOMILA. « Joint sparsity-based optimization of a set of orthonormal 2-D separable block transforms ». In : *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE. 2009, p. 9–12.
- [262] Xiaowei SUN, Baocai YIN et Yunhui SHI. « A low cost video coding scheme using texture synthesis ». In : *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*. IEEE. 2009, p. 1–5.
- [263] Po-He TSENG, Ran CARMi, Ian GM CAMERON, Douglas P MUNOZ et Laurent ITTI. « Quantifying center bias of observers in free viewing of dynamic natural scenes ». In : *Journal of vision* 9.7 (2009), p. 4.
- [264] Zhou WANG et Alan C BOVIK. « Mean squared error : love it or leave it ? A new look at signal fidelity measures ». In : *Signal Processing Magazine, IEEE* 26.1 (2009), p. 98–117.
- [265] Li-Yi WEI, Sylvain LEFEBVRE, Vivek KWATRA et Greg TURK. « State of the art in example-based texture synthesis ». In : *Eurographics 2009, State of the Art Report, EG-STAR*. Eurographics Association. 2009, p. 93–117.
- [266] Jana ZUJOVIC, Thrasyvoulos N PAPPAS et David L NEUHOFF. « Structural similarity metrics for texture analysis and retrieval ». In : *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE. 2009, p. 2225–2228.
- [267] Muhammad Zaheer AZIZ et Bärbel MERTSCHING. « Fast and robust generation of feature maps for region-based visual attention ». In : *Image Processing, IEEE Transactions on* 17.5 (2008), p. 633–644.
- [268] Marc BOSCH, Fengqing ZHU et Edward J DELP. « Video coding using motion classification ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 1588–1591.
- [269] David C BRADLEY et Manu S GOYAL. « Velocity computation in the primate visual system ». In : *Nature Reviews Neuroscience* 9.9 (2008), p. 686–695.
- [19] Olivier BROUARD, Fabrice DELANNAY, Vincent RICORDEL et Dominique BARBA. « Spatio-Temporal Segmentation and Regions Tracking of High Definition Video Sequences using a Markov Random Field Model ». In : *IEEE International Conference on Image Processing (ICIP)*. San Diego, Californie, USA. Oct. 2008.

- 
- [270] Mathieu CARNEC, Patrick LE CALLET et Dominique BARBA. « Objective quality assessment of color images based on a generic perceptual reduced reference ». In : *Signal Processing : Image Communication* 23.4 (2008), p. 239–256.
- [271] Chuo-Ling CHANG et Bernd GIROD. « Direction-adaptive partitioned block transform for image coding ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 145–148.
- [272] Roberto COSTANTINI, Luciano SBAIZ et Sabine SÜSTRUNK. « Higher order SVD analysis for dynamic texture synthesis ». In : *Image Processing, IEEE Transactions on* 17.1 (2008), p. 42–52.
- [273] Rafael G DE OLIVEIRA et Ricardo L DE QUEIROZ. « Intra prediction versus wavelets and lapped transforms in an H. 264/AVC coder ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 137–140.
- [274] Fabrice DELANNAY, Olivier BROUARD, Vincent RICORDEL et Dominique BARBA. *Lot 4.3 : Manuel d'utilisation du logiciel de conditionnement et de pré-analyse du flux vidéo*. Livrable. Projet RIAM ArchiPEG (ANR-05-RIAM-01401), juin 2008.
- [275] Fabrice DELANNAY, Olivier BROUARD, Vincent RICORDEL et Dominique BARBA. *Lot 4.4 : Rapport sur les tests des algorithmes de filtrage et de pré-analyse du flux vidéo*. Livrable. Projet RIAM ArchiPEG (ANR-05-RIAM-01401), sept. 2008.
- [276] David M HOFFMAN, Ahna R GIRSHICK, Kurt AKELEY et Martin S BANKS. « Vergence–accommodation conflicts hinder visual performance and cause visual fatigue ». In : *Journal of vision* 8.3 (2008), p. 33.
- [277] Yan HUANG, Jingliang PENG, C-C Jay KUO et M GOPI. « A generic scheme for progressive point cloud coding ». In : *IEEE Transactions on Visualization and Computer Graphics* 14.2 (2008), p. 440–453.
- [278] Guillaume JEANNIC. « Représentation structurelle d'images par transformées locales en ondelettes orientées et codage ». Thèse de doct. Université de Nantes, 2008.
- [279] Vamsi KADIYALA, Srivani PINNELI, Eric C LARSON et Damon M CHANDLER. « Quantifying the perceived interest of objects in images : effects of size, location, blur, and contrast ». In : *Electronic Imaging 2008*. ISOP. 2008, 68060S.
- [280] Aditya KHANDELIA, Saurabh GORECHA, Brejesh LALL, Santanu CHAUDHURY et Mona MATHUR. « Parametric video compression scheme using AR based texture synthesis ». In : *Computer Vision, Graphics & Image Processing, 2008. ICV-GIP'08. Sixth Indian Conference on*. IEEE. 2008, p. 219–225.
- [281] Aditya KHANDELIA, Saurabh GORECHA, Brejesh LALL, Santanu CHAUDHURY et Mona MATHUR. « Parametric video compression scheme using AR based texture synthesis ». In : *Computer Vision, Graphics & Image Processing, 2008. ICV-GIP'08. Sixth Indian Conference on*. IEEE. 2008, p. 219–225.

- [282] Kenneth KNOBLAUCH, Laurence T MALONEY et al. « MLDS : Maximum likelihood difference scaling in R ». In : *Journal of Statistical Software* 25.2 (2008), p. 1–26.
- [283] Benjamin LE GUEN. « Adaptation du contenu spatio-temporel des images pour un codage par ondelettes ». Thèse de doct. Université Rennes 1, 2008.
- [284] Khaled MAMOU, Titus ZAHARIA et Françoise PRÊTEUX. « FAMC : The MPEG-4 standard for animated mesh compression ». In : *International Conference on Image Processing*. IEEE. 2008, p. 2676–2679.
- [285] MVC. *ISO/IEC 14496-10* : 2008. URL : <http://www.iso.org/iso/>.
- [286] KS NI et Truong Q NGUYEN. « Exploiting MPEG-7 texture descriptors for fast H.264 mode decision ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 2796–2799.
- [287] Stephen E PALMER et Joseph L BROOKS. « Edge-region grouping in figure-ground organization and depth perception ». In : *Journal of Experimental Psychology : Human Perception and Performance* 34.6 (2008), p. 1353.
- [288] Sara PARRILLI, Marco CAGNAZZO et Béatrice PESQUET-POPESCU. « Distortion evaluation in transform domain for adaptive lifting schemes ». In : *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*. IEEE. 2008, p. 200–205.
- [289] Umesh RAJASHEKAR, Ian VAN DER LINDE, Alan C BOVIK et Lawrence K CORMACK. « GAFFE : A gaze-attentive fixation finding engine ». In : *Image Processing, IEEE Transactions on* 17.4 (2008), p. 564–573.
- [290] Aleksandar STOJANOVIC, Mathias WIEN et Jens-Rainer OHM. « Dynamic texture synthesis for H. 264/AVC inter coding ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 1608–1611.
- [291] Aleksandar STOJANOVIC, Mathias WIEN et J-R OHM. « Dynamic texture synthesis for H.264/AVC inter coding ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 1608–1611.
- [292] Sylvain TOURANCHEAU, Patrick LE CALLET et Dominique BARBA. « Image and video quality assessment using LCD : Comparisons with CRT conditions ». In : *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 91.6 (2008), p. 1383–1391.
- [293] Huamin WANG, Yonatan WEXLER, Eyal OFEK et Hugues HOPPE. « Factoring repeated content within and among images ». In : *ACM Transactions on Graphics (TOG)*. T. 27. 3. ACM. 2008, p. 14.
- [294] Yan YE et Marta KARCZEWICZ. « Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 2116–2119.

- [295] Bing ZENG et Jingjing FU. « Directional discrete cosine transforms—a new framework for image coding ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 18.3 (2008), p. 305–313.
- [296] Lingyun ZHANG, Matthew H TONG, Tim K MARKS, Honghao SHAN et Garrison W COTTRELL. « SUN : A Bayesian framework for saliency using natural statistics ». In : *Journal of vision* 8.7 (2008), p. 32.
- [297] Shengping ZHANG, Hongxun YAO et Shaohui LIU. « Dynamic background modeling and subtraction using spatio-temporal local binary patterns ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 1556–1559.
- [298] Xiaonan ZHAO, Matthew G REYES, Thrasyvoulos N PAPPAS et David L NEUHOFF. « Structural texture similarity metrics for retrieval applications ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 1196–1199.
- [299] Yinqing ZHAO, Krit PANUSOPONE et Limin WANG. « Fast motion estimation algorithm using unequal search effort for H.264 | MPEG-4 AVC encoder ». In : *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE. 2008, p. 1996–1999.
- [300] Chunbo ZHU, Xiaoyan SUN, Feng WU et Houqiang LI. « Video coding with spatio-temporal texture synthesis and edge-based inpainting ». In : *Multimedia and Expo, 2008 IEEE International Conference on*. 2008, p. 813–816.
- [301] Tomer AMIAZ, Sándor FAZEKAS, Dmitry CHETVERIKOV et Nahum KIRYATI. « Detecting regions of dynamic texture ». In : *Scale Space and Variational Methods in Computer Vision*. Springer, 2007, p. 848–859.
- [302] Shai AVIDAN et Ariel SHAMIR. « Seam carving for content-aware image resizing ». In : *ACM Transactions on graphics (TOG)*. T. 26. 3. ACM. 2007, p. 10.
- [303] Johannes BALLÉ et Mathias WIEN. « Extended texture prediction for H.264/AVC intra coding ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 6. IEEE. 2007, p. VI–93.
- [304] Hugues BALLUET. *Manuel de conception de la plate-forme matérielle - Lot 2.2*. Rapp. tech. Projet RIAM ArchiPEG (ANR-05-RIAM-01401), juin 2007.
- [305] Yukihiro BANDO, Kazuya HAYASE, Seishi TAKAMURA, Kazuto KAMIKURA et Yoshiyuki YASHIMA. « Mode Decision for H. 264/AVC based on Spatio-Temporal Sensitivity ». In : *Proceedings of the Picture Coding Symposium, PCS*. Citeseer. 2007.
- [306] Aina BARCELO, Eduard MONTSENY et Pilar SOBREVILLA. « Fuzzy texture unit and fuzzy texture spectrum for texture characterization ». In : *Fuzzy Sets and Systems* 158.3 (2007), p. 239–252.



- [307] Marc BOSCH, Fengqing ZHU et Edward J DELP. « Spatial texture models for video compression ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 1. IEEE. 2007, p. I–93.
- [40] Olivier BROUARD, Fabrice DELANNAY, Vincent RICORDEL et Dominique BARBA. « Estimation Robuste du Mouvement Global pour des Séquences Vidéo Haute Définition en utilisant une Estimation Multi-Résolutions du Mouvement ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Montpellier, France. Nov. 2007.
- [20] Olivier BROUARD, Fabrice DELANNAY, Vincent RICORDEL et Dominique BARBA. « Fast Long-Term Motion Estimation for High Definition Video Sequences Based on Spatio-Temporal Tubes and using the Nelder-Mead Simplex Algorithm ». In : *Picture Coding Symposium (PCS)*. Lisbonne, Portugal. Nov. 2007.
- [21] Olivier BROUARD, Fabrice DELANNAY, Vincent RICORDEL et Dominique BARBA. « Robust Motion Segmentation for High Definition Video Sequences using a Fast Multi-Resolution Motion Estimation based on Spatio-Temporal Tubes ». In : *Picture Coding Symposium (PCS)*. Lisbonne, Portugal. Nov. 2007.
- [308] Alexandre BUR et Heinz HÜGLI. « Optimal cue combination for saliency computation : A comparison with human vision ». In : *Nature Inspired Problem-Solving Methods in Knowledge Engineering*. Springer, 2007, p. 109–118.
- [309] Damon M CHANDLER et Sheila S HEMAMI. « VSNR : A wavelet-based visual signal-to-noise ratio for natural images ». In : *Image Processing, IEEE Transactions on* 16.9 (2007), p. 2284–2298.
- [310] Chuo-Ling CHANG et Bernd GIROD. « Direction-adaptive discrete wavelet transform for image compression ». In : *Image Processing, IEEE Transactions on* 16.5 (2007), p. 1289–1302.
- [311] Christophe CHARRIER, Laurence T MALONEY, Hocine CHERIFI et Kenneth KNOBLAUCH. « Maximum likelihood difference scaling of image quality in compression-degraded images ». In : *JOSA A* 24.11 (2007), p. 3418–3426.
- [312] Fabrice DELANNAY, Olivier BROUARD, Vincent RICORDEL et Dominique BARBA. *Lot 4.2 : Définitions d'algorithmes de filtrage et de pré-analyse du flux vidéo*. Livrable. Projet RIAM ArchiPEG (ANR05RIAM01401), avr. 2007.
- [313] Wenpeng DING, Feng WU, Xiaolin WU, Shipeng LI et Houqiang LI. « Adaptive directional lifting-based wavelet transform for image coding ». In : *Image Processing, IEEE Transactions on* 16.2 (2007), p. 416–427.
- [314] Video Quality Experts GROUP. *VQEG multimedia test plan 1.19*. 2007.
- [315] Olivier GUYE. *Etat de l'art sur les architectures de calcul hautes performances - Lot 2.1*. Rapp. tech. Projet RIAM ArchiPEG (ANR-05-RIAM-01401), juin 2007.
- [316] Xiaodi HOU et Liqing ZHANG. « Saliency detection : A spectral residual approach ». In : *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE. 2007, p. 1–8.

- 
- [317] Chiuan HWANG, ShinShan ZHUANG et Shang-Hong LAI. « Efficient intra mode selection using image structure tensor for H.264/AVC ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 5. IEEE. 2007, p. V–289.
- [22] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « A Multiresolution Approach for the Coding of Edges of Still Images using Adaptive Arithmetic Coding ». In : *Picture Coding Symposium (PCS)*. Lisbonne, Portugal. Nov. 2007.
- [41] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « Représentation structurelle d’images fixes par transformée en ondelettes orientées ». In : *Colloque GRETSI sur le traitement du signal et des images (GRETSI)*. Troyes, France. Sept. 2007.
- [23] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « The Edge Driven Oriented Wavelet Transform : An Anisotropic Multidirectional Representation With Oriented Lifting Scheme ». In : *Visual Communications and Image Processing (VCIP) part of the SPIE Symposium on Electronic Imaging*. San Jose, Californie, USA. Jan. 2007.
- [318] Spyridon K KAPOTAS et Athanassios N SKODRAS. « Fast multiple reference frame selection method in H.264 video encoding ». In : *Proceedings of the Picture Coding Symposium, PCS*. Citeseer. 2007.
- [319] Byung-Gyu KIM et Chang-Sik CHO. « A fast inter-mode decision algorithm based on macro-Block tracking for P slices in the H.264/AVC video standard ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 5. IEEE. 2007, p. V–301.
- [320] Byeongdu LA, Minyoung EOM et Yoonsik CHOE. « Fast mode decision for intra prediction in H.264/AVC encoder ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 5. IEEE. 2007, p. V–321.
- [321] Olivier LE MEUR, Patrick LE CALLET et Dominique BARBA. « Predicting visual fixations on video based on low-level visual features ». In : *Vision research* 47.19 (2007), p. 2483–2498.
- [322] Yu-Ming LEE et Yinyi LIN. « An improved zero-block mode decision algorithm for H.264/AVC ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 5. IEEE. 2007, p. V–293.
- [323] Dong LIU, Xiaoyan SUN, Feng WU, Shipeng LI et Ya-Qin ZHANG. « Image compression with edge-based inpainting ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 17.10 (2007), p. 1273–1287.
- [324] Laurens van der MAATEN et Eric POSTMA. « Texton-based texture classification ». In : *Proceedings of Belgium-Netherlands Artificial Intelligence Conference*. 2007.

- [325] Eduardo MARTINEZ-ENRIQUEZ, Manuel DE-FRUTOS-LOPEZ, Jose Carlos PUJOL-ALCOLADO et Fernando DIAZ-DE-MARIA. « A fast motion-cost based algorithm for H.264/AVC inter mode decision ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 5. IEEE. 2007, p. V–325.
- [326] Javier A MONTOYA-ZEGARRA, Neucimar Jerônimo LEITE et R da S TORRES. « Rotation-invariant and scale-invariant steerable pyramid decomposition for texture image retrieval ». In : *Computer Graphics and Image Processing, 2007. SIBGRAPI 2007. XX Brazilian Symposium on*. IEEE. 2007, p. 121–128.
- [327] Patrick NDJIKI-NYA, Tobias HINZ et Thomas WIEGAND. « Generic and robust video coding with texture analysis and synthesis ». In : *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE. 2007, p. 1447–1450.
- [328] Yu-Nan PAN et Tsung-Han TSAI. « Fast motion estimation and edge information inter-mode decision on H.264 video coding ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 2. IEEE. 2007, p. II–473.
- [329] Stéphane PÉCHARD, Patrick LE CALLET, Mathieu CARNEC et Dominique BARBA. « A new methodology to estimate the impact of H. 264 artefacts on subjective video quality ». In : *International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM*. 2007, p. 373.
- [330] Teodora PETRISOR, Béatrice PESQUET-POPESCU et Jean-Christophe PESQUET. « A compressed sensing approach to frame-based multiple description coding ». In : *ICASSP*. 2007, p. 709–712.
- [331] Daniel SCHARSTEIN et Chris PAL. « Learning conditional random fields for stereo ». In : *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE. 2007, p. 1–8.
- [332] Chang SUN, Hong-Jun WANG, Hua LI et Tai-hoon KIM. « Perceptually adaptive Lagrange multiplier for rate-distortion optimization in H.264 ». In : *Future Generation Communication and Networking (FGCN 2007)*. T. 1. IEEE. 2007, p. 459–463.
- [333] Benjamin W TATLER. « The central fixation bias in scene viewing : Selecting an optimal viewing position independently of motor biases and image feature distributions ». In : *Journal of Vision* 7.14 (2007), p. 4.
- [334] Dajun WU, Keng Pang LIM, Tuan Kiang CHIEW, Jo Yew THAM et KH GOH. « An adaptive thresholding technique for the detection of all-zeros blocks in H.264 ». In : *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. T. 5. IEEE. 2007, p. V–329.
- [335] Kai ZHANG, Qiang WANG, Qian HUANG, Debin ZHAO et Wen GAO. « A Context-based Adaptive Fast Intra\_4x4 Prediction Mode Decision Algorithm for H.264/AVC Video Coding ». In : *PCS, Nov*. 2007.

- [336] Guoying ZHAO et Matti PIETIKAINEN. « Dynamic texture recognition using local binary patterns with an application to facial expressions ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.6 (2007), p. 915–928.
- [337] HK ATES, Berkay KANBEROGLU et Yucel ALTUNBASAK. « Rate-distortion and complexity joint optimization for fast motion estimation in H.264 video coding ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 37–40.
- [338] Siley BA, Yucel ALTUNBASAK et Hasan ATES. « Low complexity inter-mode selection for H.264 ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 1349–1352.
- [339] Olivier BROUARD. *Lot 4.1 : Étude sur les méthodes de conditionnement et de pré-analyse du flux vidéo*. Livrable. Projet RIAM ArchiPEG (ANR05RIAM01401), sept. 2006.
- [340] Jiajun BU, Shuiyong LOU, Chun CHEN et Jingjing ZHU. « A predictive block-size mode selection for inter frame in H.264 ». In : *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. T. 2. IEEE. 2006.
- [341] Wei-Han CHANG, Nai-Chung YANG, Chung-Ming KUO, Yi-Jhen CHEN et al. « An efficient temporal texture descriptor for video retrieval ». In : *Proceedings of the 6th WSEAS International Conference on Signal Processing, Computational Geometry & Artificial Vision*. World Scientific, Engineering Academy et Society (WSEAS). 2006, p. 107–112.
- [342] Vivien CHAPPELIER et Christine GUILLEMOT. « Oriented wavelet transform for image compression and denoising ». In : *Image Processing, IEEE Transactions on* 15.10 (2006), p. 2892–2903.
- [343] Arthur L DA CUNHA, Jianping ZHOU et Minh N DO. « The Nonsubsampled Contourlet Transform : Theory, Design and Applications ». In : *Image Processing, IEEE Transactions on* 15.10 (2006), p. 3089–3101.
- [344] Jun-Ren DING, Jar-Ferr YANG et Ji-Kun LIN. « Motion-based Adaptive GOP Algorithms for Efficient H. 264/AVC Compression ». In : *JCIS*. 2006.
- [345] Paul R HILL, Tuan-Kiang CHIEW et David R BULL. « Interpolation free sub-pixel motion estimation for H.264 ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 1337–1340.
- [346] Win-Bin HUANG, Yi-Li LIN, Hung-Wei CHENG, Alvin WY SU et Yau-Hwang KUO. « Two-Stage Mode Selection of H.264/AVC Video Encoding with Rate Distortion Optimization ». In : *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. T. 2. IEEE. 2006.

- [42] Guillaume JEANNIC, Vincent RICORDEL et Dominique BARBA. « Extraction de contours multirésolution pour un codage d'images par bandelettes ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Caen, France. Nov. 2006.
- [347] Liangbao JIAO, De ZHANG et Houjie BI. « Inherit-based adaptive frame selection for fast multi-frame motion estimation in H. 264 ». In : *Intelligent Computing in Signal Processing and Pattern Recognition*. Springer, 2006, p. 938–944.
- [348] Zoltan KATO et Ting-Chuen PONG. « A Markov random field image segmentation model for color textured images ». In : *Image and Vision Computing 24.10* (2006), p. 1103–1114.
- [349] Byung-Gyu KIM, Suk-Kyu SONG et Chang-Sik CHO. « Efficient inter-mode decision based on contextual prediction for the P-slice in H.264/AVC video coding ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 1333–1336.
- [350] Chong-Kwon KIM, Siwei MA et CJ KUO. « Fast H.264 motion estimation with block-size adaptive referencing (BAR) ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 57–60.
- [351] Olivier LE MEUR, Patrick LE CALLET, Dominique BARBA et Dominique THOREAU. « A coherent computational approach to model bottom-up visual attention ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on 28.5* (2006), p. 802–817.
- [352] Pei-Jun LEE et Ming-Long LIN. « Fast inter mode selection algorithm for motion estimation in MPEG-4 AVC/JVT/H.264 ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 1365–1368.
- [353] Qiong LIU, Rui-min HU, LI ZHU, Xin-chen ZHANG et Zhen HAN. « Improved fast intra prediction algorithm of H.264/AVC ». In : *Journal of Zhejiang University Science A 7.1* (2006), p. 101–105.
- [354] Zhenyu LIU, Yang SONG, Takeshi IKENAGA et Satoshi GOTO. « Low-pass filter based vlsi oriented variable block size motion estimation algorithm for H.264 ». In : *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. T. 2. IEEE. 2006, p. II–II.
- [355] Zhi-Yi MAI, Chun-Ling YANG, Kai-Zhi KUANG et Lai-Man PO. « A novel motion estimation method based on structural similarity for H.264 inter prediction ». In : *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. T. 2. IEEE. 2006.
- [356] Patrick NDJIKI-NYA, Tobias HINZ, Christoph STUBER et Thomas WIEGAND. « A content-based video coding approach for rigid and non-rigid textures ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 3169–3172.

- 
- [357] Marcos NIETO, Luis SALGADO et Julian CABRERA. « Fast mode decision on H.264/AVC main profile encoding based on PSNR predictions ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 49–52.
- [358] Nicole C RUST, Valerio MANTE, Eero P SIMONCELLI et J Anthony MOVSHON. « How MT cells analyze the motion of visual patterns ». In : *Nature neuroscience* 9.11 (2006), p. 1421–1431.
- [359] Luis SALGADO et Marcos NIETO. « Sequence independent very fast mode decision algorithm on H.264/AVC baseline profile ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 41–44.
- [360] Ruwen SCHNABEL et Reinhard KLEIN. « Octree-based Point-Cloud Compression. » In : *Eurographics / IEEE VGTC Conference on Point-Based Graphics SPBG 6* (2006), p. 111–120.
- [361] H. R. SHEIKH, M. F. SABIR et A. C. BOVIK. « A statistical evaluation of recent full reference image quality assessment algorithms ». In : *Image Processing, IEEE Transactions on* 15.11 (2006).
- [362] Hamid Rahim SHEIKH et Alan C BOVIK. « Image information and visual quality ». In : *Image Processing, IEEE Transactions on* 15.2 (2006), p. 430–444.
- [363] Aljoscha SMOLIC, Karsten MUELLER, Philipp MERKLE, Christoph FEHN, Peter KAUFF, Peter EISERT et Thomas WIEGAND. « 3D video and free viewpoint video-technologies, applications and MPEG standards ». In : *Multimedia and Expo, 2006 IEEE International Conference on*. Toronto, Canada. 2006, p. 2161–2164.
- [364] Rui SU, Guizhong LIU et Tongyu ZHANG. « Fast mode decision algorithm for intra prediction in H.264/AVC ». In : *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. T. 2. IEEE. 2006, p. II–II.
- [365] Yeping SU et Ming-Ting SUN. « Fast multiple reference frame motion estimation for H.264/AVC ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 16.3 (2006), p. 447–452.
- [366] Tsung-Han TSAI et Yu-Nan PAN. « A novel 3-D predict hexagon search algorithm for fast block motion estimation on H. 264 video coding ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 16.12 (2006), p. 1542–1549.
- [367] Kazuhiko UKAI. « Human factors for ftereoscopic images ». In : *International Conference on Multimedia and Expo*. IEEE. 2006, p. 1697–1700.
- [368] Vladan VELISAVLJEVIC, Baltasar BEFERULL-LOZANO, Martin VETTERLI et Pier Luigi DRAGOTTI. « Directionlets : anisotropic multidirectional representation with separable filtering ». In : *Image Processing, IEEE Transactions on* 15.7 (2006), p. 1916–1933.
- [369] Chou-Chen WANG, Tsung-Shien CHEN, Chi-Wei TUNG et al. « Fast Intra-Mode Decision in H.264 using Interblock Correlation ». In : *ICIP*. 2006, p. 1345–1348.

- [370] Demin WANG, Liang ZHANG, André VINCENT et Filippo SPERANZA. « Curved wavelet transform for image coding ». In : *Image Processing, IEEE Transactions on* 15.8 (2006), p. 2413–2421.
- [371] Hanli WANG, Sam KWONG et Chi-Wah KOK. « Effectively detecting all-zero DCT blocks for H. 264 optimization ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 1329–1332.
- [372] Zhou WANG et Alan C BOVIK. « Modern image quality assessment ». In : *Synthesis Lectures on Image, Video, and Multimedia Processing 2.1* (2006), p. 1–156.
- [373] Ying-Lai XI, Chong-Yang HAO, Yang-Yu FAN et Hong-Qi HU. « A fast block-matching algorithm based on adaptive search area and its VLSI architecture for H. 264/AVC ». In : *Signal processing : Image communication* 21.8 (2006), p. 626–646.
- [374] Jun XIN et Anthony VETRO. « Fast mode decision for intra-only H.264/AVC coding ». In : *Picture Coding Symposium (PCS)*. 2006.
- [375] Yiwen ZHAO, Maja BYSTROM et Ita RICHARDSON. « A MAP framework for efficient skip/code mode decision in H.264 ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 45–48.
- [376] Bobby ABRAHAM, Octavia I CAMPS et Mario SZNAIER. « Dynamic texture with Fourier descriptors ». In : *Proceedings of the 4th International Workshop on Texture Analysis and Synthesis*. 2005, p. 53–58.
- [377] Laurent AIMAR, Loren MERRITT, Eric PETIT, Min CHEN, Justin CLAY, Mns RULLGRD, Christian HEINE et Alex IZVORSKI. « x264-a free h264/AVC encoder ». In : *Online (last accessed on : 04/01/07) : <http://www.videolan.org/developers/x264.html>* (2005).
- [378] Antoni B CHAN et Nuno VASCONCELOS. « Probabilistic kernels for the classification of auto-regressive visual processes ». In : *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. T. 1. IEEE. 2005, p. 846–851.
- [379] Chun CHEN, Linjian MO, Jiajun BU, Shuiyong LOU et Zhi YANG. « A novel fast predictive mode decision algorithm for H.264 ». In : *Signal Processing and Its Applications, 2005. Proceedings of the Eighth International Symposium on*. T. 1. IEEE. 2005, p. 243–246.
- [380] Dmitry CHETVERIKOV et Renaud PÉTERI. « A brief survey of dynamic texture description and recognition ». In : *Computer Recognition Systems*. Springer, 2005, p. 17–26.
- [381] Byeong-Doo CHOI, Min-Cheol HWANG, Jun-Ki CHO, Jin-Sam KIM, Jin-Hyung KIM et Sung-Jea KO. « Realtime H. 264 encoding system using fast motion estimation and mode decision ». In : *Embedded and Ubiquitous Computing-EUC 2005*. Springer, 2005, p. 174–183.

- [382] Minh N DO et Martin VETTERLI. « The contourlet transform : an efficient directional multiresolution image representation ». In : *Image Processing, IEEE Transactions on* 14.12 (2005), p. 2091–2106.
- [383] Michael ELAD, J-L STARCK, Philippe QUERRE et David L DONOHO. « Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA) ». In : *Applied and Computational Harmonic Analysis* 19.3 (2005), p. 340–358.
- [384] Min-Cheol HWANG, Jun-Ki CHO, Jin-Sam KIM, Jin-Hyung KIM et Sung-Jea KO. « Fast intra prediction mode selection scheme using temporal correlation in H.264 ». In : *TENCON 2005 2005 IEEE Region 10*. IEEE. 2005, p. 1–5.
- [385] Geun-Yong KIM, Seung-Hwan KIM, Hee-Soon KIM et Yo-Sung HO. « Fast mode decision algorithm for H.264 based on motion cost ». In : *Signal Processing Conference, 2005 13th European*. IEEE. 2005, p. 1–4.
- [386] Sanjeev KUMAR, Mainak BISWAS, Serge J BELONGIE et Truong Q NGUYEN. « Spatio-temporal texture synthesis and image painting for video applications ». In : *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. T. 2. 85. 2005.
- [387] Vivek KWATRA, Irfan ESSA, Aaron BOBICK et Nipun KWATRA. « Texture optimization for example-based synthesis ». In : *ACM Transactions on Graphics (TOG)*. T. 24. 3. ACM. 2005, p. 795–802.
- [388] Olivier LE MEUR. « Attention sélective en visualisation d’images fixes et animées affichées sur écran : modèles et évaluation de performances-application ». Thèse de doct. Université de Nantes, 2005.
- [389] Olivier LE MEUR, Dominique BARBA, Patrick LE CALLET et Dominique THOREAU. « A human visual model-based approach of the visual attention and performance evaluation ». In : *Electronic Imaging*. International Society for Optics et Photonics. 2005, p. 258–267.
- [390] Olivier LE MEUR, Dominique THOREAU, Patrick LE CALLET et Dominique BARBA. « A spatio-temporal model of the selective human visual attention ». In : *Image Processing (ICIP) International Conference on*. T. 3. IEEE. 2005, p. 1188.
- [391] Erwan LE PENNEC et Stéphane MALLAT. « Sparse geometric image representations with bandelets ». In : *Image Processing, IEEE Transactions on* 14.4 (2005), p. 423–438.
- [392] Patrick NDJIKI-NYA, Tobias HINZ, Aljoscha SMOLIC et Thomas WIEGAND. « A generic and automatic content-based approach for improved H.264/MPEG4-AVC video coding ». In : *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. T. 2. IEEE. 2005, p. II-874.



- [393] Robert J O'CALLAGHAN et David R BULL. « Combined morphological-spectral unsupervised image segmentation ». In : *Image Processing, IEEE Transactions on* 14.1 (2005), p. 49–62.
- [394] JM ODOBEZ, P BOUTHEMY et F SPINDLER. *Motion2D : a software to estimate 2D parametric motion models*. <http://www.irisa.fr/Vista/Motion2D/>. 2005.
- [395] Jens-Rainer OHM. « Advances in scalable video coding ». In : *Proceedings of the IEEE* 93.1 (2005), p. 42–56.
- [396] Feng PAN, Xiao LIN, Susanto RAHARDJA, Keng Pang LIM, ZG LI, Dajun WU et Si WU. « Fast mode decision algorithm for intraprediction in H. 264/AVC video coding ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 15.7 (2005), p. 813–822.
- [397] Renaud PÉTERI et Dmitry CHETVERIKOV. « Dynamic texture recognition using normal flow and texture regularity ». In : *Pattern Recognition and Image Analysis*. Springer, 2005, p. 223–230.
- [398] Branko PETLJANSKI et Oge MARQUES. « A Novel Approach for Video Quantization Using the Spatiotemporal Frequency Characteristics of the Human Visual System ». In : *BMVC*. 2005.
- [399] Gabriel PEYRÉ et Stéphane MALLAT. « Discrete bandelets with geometric orthogonal filters ». In : *IEEE International Conference on Image Processing, 2005. (ICIP 2005)*. T. 1. IEEE. 2005, p. 65–68.
- [400] Gabriel PEYRÉ et Stéphane MALLAT. « Surface compression with geometric bandelets ». In : *ACM Transactions on Graphics (TOG)* 24.3 (2005), p. 601–608.
- [401] Ashfaqur RAHMAN et Manzur MURSHED. « A motion-based approach for temporal texture synthesis ». In : *TENCON 2005 2005 IEEE Region 10*. IEEE. 2005, p. 1–4.
- [35] Vincent RICORDEL, Florent AUTRUSSEAU, William DUPUY et Dominique BARBA. « 1D-Mosaics Grouping using Lattice Vector Quantization for a Video Browsing Application ». In : *International Workshop on Content-Based Multimedia Indexing (CBMI)*. Riga, Lettonie. Juin 2005.
- [24] Vincent RICORDEL, Patrick Le CALLET, Mathieu CARNEC et Dominique BARBA. « Flexible Storage of Still Images with a Perceptual Quality Criterion ». In : *Advanced Concepts for Intelligent Vision Systems (ACIVS)*. Anvers, Belgique. Sept. 2005.
- [402] Hamid R SHEIKH, Zhou WANG, Lawrence CORMACK et Alan C BOVIK. *LIVE image quality assessment database release 2*. 2005.
- [403] Thoma SIKORA. « Trends and perspectives in image and video coding ». In : *Proceedings of the IEEE* 93.1 (2005), p. 6–17.
- [404] P. SUNNA. « AVC/H.264 – an advanced video coding system for SD and HD broadcasting ». In : *EBU Technical Review* (avr. 2005).

- [405] Tatsuto TAKEUCHI et Karen K DE VALOIS. « Sharpening image motion based on the spatio-temporal characteristics of human vision ». In : *Electronic Imaging 2005*. International Society for Optics et Photonics. 2005, p. 83–94.
- [406] Alexis Michael TOURAPIS, Hye-Yeon CHEONG et Pankaj TOPIWALA. « Fast ME in the JM reference software ». In : *JVT Document JVT-P026, 16th Meeting*. 2005, p. 24–29.
- [407] Takeshi TSUKUBA, Isao NAGAYOSHI, Tsuyoshi HANAMURA et Hideyoshi TOMINAGA. « H.264 fast intra-prediction mode decision based on frequency characteristic ». In : *Signal Processing Conference, 2005 13th European*. IEEE. 2005, p. 1–4.
- [408] Yu-Lin WANG, Jing-Xin WANG, Yen-Wen LAI et Alvin WY SU. « Dynamic GOP structure determination for real-time MPEG-4 advanced simple profile video encoder ». In : *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE. 2005, p. 293–296.
- [409] Simon J WATT, Kurt AKELEY, Marc O ERNST et Martin S BANKS. « Focus cues affect perceived depth ». In : *Journal of Vision* 5.10 (2005), p. 7.
- [410] Xiaoquan YI, Jun ZHANG, Nam LING et Weijia SHANG. « Improved and simplified fast motion estimation for JM ». In : *JVT-P021. doc, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 16th meeting, Pozan, Poland*. 2005.
- [411] Andy C YU, Graham R MARTIN et Heechan PARK. « Improved schemes for inter-frame coding in the H.264/AVC standard ». In : *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. T. 2. IEEE. 2005, p. II–902.
- [412] Hongtao YU, Feng PAN, Zhiping LIN et Yin SUN. « A perceptual bit allocation scheme for H.264 ». In : *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE. 2005, 4–pp.
- [413] Geert CAENEN et Luc VAN GOOL. « Maximum Response Filters for Texture Analysis ». In : *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*. IEEE. 2004, p. 58–58.
- [414] Neill CAMPBELL, Colin DALTON, David GIBSON, David OZIEM et Barry THOMAS. « Practical generation of video textures using the auto-regressive process ». In : *Image and Vision Computing* 22.10 (2004), p. 819–827.
- [415] Che-Yu CHANG, Chia-Ho PAN et Homer CHEN. « Fast mode decision for P-frames in H.264 ». In : *Picture Coding Symposium, San Francisco, CA*. 2004.
- [416] Mei-Juan CHEN, Yi-Yen CHIANG, Hung-Ju LI et Ming-Chieh CHI. « Efficient multi-frame motion estimation algorithms for MPEG-4 AVC/JVT/H.264 ». In : *Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on*. T. 3. IEEE. 2004, p. III–737.
- [417] Renan COUDRAY et Bernard BESSERER. « Global motion estimation for MPEG-encoded streams ». In : *Image Processing, 2004. ICIP'04. 2004 International Conference on*. T. 5. IEEE. 2004, p. 3411–3414.

- [418] Stephen V DAVID, William E VINJE et Jack L GALLANT. « Natural stimulus statistics alter the receptive field structure of V1 neurons ». In : *The Journal of Neuroscience* 24.31 (2004), p. 6991–7006.
- [419] Adriana DUMITRAS et Barry G HASKELL. « I/P/B frame type decision by collinearity of displacements ». In : *Image Processing, 2004. ICIP'04. 2004 International Conference on*. T. 4. IEEE. 2004, p. 2769–2772.
- [420] Laurent ITTI. « Automatic foveation for video compression using a neurobiological model of visual attention ». In : *Image Processing, IEEE Transactions on* 13.10 (2004), p. 1304–1318.
- [421] John IVE. « Images formats for HDTV ». In : *EBU Technical Review* (juil. 2004).
- [422] Michael S LANDY et Norma GRAHAM. « Visual perception of texture ». In : *The visual neurosciences 2* (2004), p. 1106–1118.
- [423] Olivier LE MEUR, Patrick LE CALLET, Dominique BARBA et Dominique THOREAU. « Performance assessment of a visual attention system entirely based on a human vision modeling ». In : *Image Processing (ICIP) International Conference on*. T. 4. IEEE. 2004, p. 2327–2330.
- [424] Olivier LE MEUR, Patrick LE CALLET, Dominique BARBA, Dominique THOREAU et Edouard FRANCOIS. « From low-level perception to high-level perception : a coherent approach for visual attention modeling ». In : *Electronic Imaging 2004. International Society for Optics et Photonics*. 2004, p. 284–295.
- [425] Xiang LI, Xiang LI et Yen-Kuang CHEN. « Fast multi-frame motion estimation algorithm with adaptive search strategies in H.264 ». In : *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*. T. 3. IEEE. 2004, p. III–369.
- [426] Lydia MJ MEESTERS, Wijnand ISSSELSTEIJN, Pieter JH SEUNTIËNS et al. « A survey of perceptual evaluations and requirements of three-dimensional TV ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 14.3 (2004), p. 381–391.
- [427] Jörg ONTRUP, Heiko WERSING et Helge RITTER. « A computational feature binding model of human texture perception ». In : *Cognitive Processing* 5.1 (2004), p. 31–44.
- [428] Jörn OSTERMANN, Jan BORMANS, Peter LIST, Detlev MARPE, Matthias NARRO-SCHKE, Fernando PEREIRA, Thomas STOCKHAMMER et Thomas WEDI. « Video coding with H. 264/AVC : tools, performance, and complexity ». In : *Circuits and Systems magazine, IEEE* 4.1 (2004), p. 7–28.
- [429] Derrick J PARKHURST et Ernst NIEBUR. « Texture contrast attracts overt visual attention in natural scenes ». In : *European Journal of Neuroscience* 19.3 (2004), p. 783–789.
- [430] John A PERRONE. « A visual motion sensor based on the properties of V1 and MT neurons ». In : *Vision research* 44.15 (2004), p. 1733–1755.

- 
- [431] Fatih PORIKLI et Yao WANG. « Automatic video object segmentation using volume growing and hierarchical clustering ». In : *EURASIP Journal on Applied Signal Processing* (2004), p. 814–832.
- [432] Ashfaqur RAHMAN et Manzur MURSHED. « Real-time temporal texture characterisation using block-based motion co-occurrence statistics ». In : *International Conference on Image Processing*. 2004.
- [433] Yeping SU et Ming-Ting SUN. « Fast multiple reference frame motion estimation for H.264 ». In : *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*. T. 1. IEEE. 2004, p. 695–698.
- [434] ITU International Telecommunication UNION. *Methodology for the subjective assessment of the quality of television pictures*. Recommendation ITU-R BT.500-11. 2004.
- [36] Pierre VERBERT, Vincent RICORDEL et Jeanpierre GUÉDON. « Analysis of Mojette Transform Projections for an efficient Coding ». In : *Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*. Lisbonne, Portugal. Avr. 2004.
- [435] Z. WANG, A. C. BOVIK, H. R. SHEIKH et E. P. SIMONCELLI. « Image quality assessment : from error visibility to structural similarity ». In : *IEEE Transactions on Image Processing* 13.4 (2004), p. 600–612.
- [436] Jeremy M WOLFE et Todd S HOROWITZ. « What attributes guide the deployment of visual attention and how do they do it? » In : *Nature Reviews Neuroscience* 5.6 (2004), p. 495–501.
- [437] Andy C YU. « Efficient block-size selection algorithm for inter-frame coding in H.264/MPEG-4 AVC ». In : *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*. T. 3. IEEE. 2004, p. III–169.
- [438] Andy C YU et Graham R MARTIN. « Advanced block size selection algorithm for inter frame coding in H.264/MPEG-4 AVC ». In : *Image Processing, 2004. ICIP'04. 2004 International Conference on*. T. 1. IEEE. 2004, p. 95–98.
- [439] Lu YUAN, Fang WEN, Ce LIU et Heung-Yeung SHUM. « Synthesizing dynamic texture with closed-loop linear dynamic system ». In : *Computer Vision-ECCV 2004*. Springer, 2004, p. 603–616.
- [440] Yu YUAN, David FENG et Yu-Zhuo ZHONG. « Three Fast Methods for Adaptive Key-frame Setting and Dynamic Frame-rate Adjusting in Video Coding ». In : *International Journal of Computational Intelligence* 1.1 (2004).
- [441] Peter GJ BARTEN. « Formula for the contrast sensitivity of the human eye ». In : *Electronic Imaging*. International Society for Optics et Photonics. 2003, p. 231–238.

- [442] Marcelo BERTALMIO, Luminita VESE, Guillermo SAPIRO et Stanley OSHER. « Simultaneous structure and texture image inpainting ». In : *Image Processing, IEEE Transactions on* 12.8 (2003), p. 882–889.
- [443] Andrew P BRADLEY. « Can region of interest coding improve overall perceived image quality ? » In : *APRS Workshop on Digital Image Computing (WDIC'03)*. T. 1. The University of Queensland. 2003, p. 41–44.
- [444] Neil DB BRUCE et M JERNIGAN. « Evolutionary design of context-free attentional operators ». In : *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*. T. 1. IEEE. 2003, p. I–429.
- [445] Roxanne CANOSA. « Seeing, sensing, and selection : modeling visual perception in complex environments ». Thèse de doct. Rochester Institute of Technology, 2003.
- [446] Andy CHANG, OC AN et Yick Ming YEUNG. « A novel approach to fast multi-frame selection for H.264 video coding ». In : *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*. T. 3. IEEE. 2003, p. III–413.
- [447] Marc CHAUMONT. « Représentation en objets vidéo pour un codage progressif et concurrentiel des séquences d'images ». Thèse de doct. Université Rennes 1, 2003.
- [448] Alan CHAUVIN. « Perception des scènes naturelles : étude et simulation du rôle de l'amplitude, de la phase et de la saillance dans la catégorisation et l'exploration de scènes naturelles ». Thèse de doct. Grenoble 1, 2003.
- [449] Zhibo CHEN, JianFeng XU, Yun HE et GuoZhong WANG. « Simplifications on fast motion estimation ». In : *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q. 6, document JVT-H026), 8ème meeting : Genève, Suisse*. 2003.
- [450] Wei DAI et Trac D TRAN. « Regularity-constrained pre-and post-filtering for block DCT-based systems ». In : *Signal Processing, IEEE Transactions on* 51.10 (2003), p. 2568–2581.
- [451] Nitin DHAVALÉ et Laurent ITTI. « Saliency-based multifoveated MPEG compression ». In : *Signal processing and its applications, 2003. Proceedings. Seventh international symposium on*. T. 1. IEEE. 2003, p. 229–232.
- [452] Gianfranco DORETTO, Alessandro CHIUSO, Ying Nian WU et Stefano SOATTO. « Dynamic textures ». In : *International Journal of Computer Vision* 51.2 (2003), p. 91–109.
- [453] Gianfranco DORETTO, Alessandro CHIUSO, Ying Nian WU et Stefano SOATTO. « Dynamic textures ». In : *International Journal of Computer Vision* 51.2 (2003), p. 91–109.

- [454] Gianfranco DORETTO et Stefano SOATTO. « Editable dynamic textures ». In : *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. T. 2. IEEE. 2003, p. II–137.
- [455] Adriana DUMITRAS et Barry G HASKELL. « A texture replacement method at the encoder for bit-rate reduction of compressed video ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 13.2 (2003), p. 163–175.
- [456] Guoliang FAN et Xiang-Gen XIA. « Wavelet-based texture analysis and synthesis using hidden Markov models ». In : *Circuits and Systems I : Fundamental Theory and Applications, IEEE Transactions on* 50.1 (2003), p. 106–120.
- [457] Xiaodong GU et Hongjiang ZHANG. « Implementing dynamic GOP in video encoding ». In : *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*. T. 1. IEEE. 2003, p. I–349.
- [458] Yu-Wen HUANG, Bing-Yu HSIEH, Tu-Chih WANG, Shao-Yi CHEN, Shyh-Eh MA, Chun-Fu SHEN et Liang-Gee CHEN. « Analysis and reduction of reference frames for motion estimation in MPEG-4 AVC/JVT/H.264 ». In : *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*. T. 2. IEEE. 2003, p. II–809.
- [459] Vivek KWATRA, Arno SCHÖDL, Irfan ESSA, Greg TURK et Aaron BOBICK. « Graphcut textures : image and video synthesis using graph cuts ». In : *ACM Transactions on Graphics (ToG)*. T. 22. 3. ACM. 2003, p. 277–286.
- [460] KP LIM, S WU, DJ WU, S RAHARDJA, X LIN, F PAN et ZG LI. « Fast inter mode selection ». In : *JVT-I020, JVT 9th Meeting, san Diego, USA*. 2003.
- [461] Kai-Kuang MA et Gang QIU. « Unequal-arm adaptive rood pattern search for fast block-matching motion estimation in the JVT/H.26L ». In : *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*. T. 1. IEEE. 2003, p. I–901.
- [462] Laurence T MALONEY et Joong Nam YANG. « Maximum likelihood difference scaling ». In : *Journal of Vision* 3.8 (2003), p. 5.
- [463] Patrick NDJIKI-NYA, Bela MAKAI, Gabi BLATTERMANN, Aljoscha SMOLIC, Heiko SCHWARZ et Thomas WIEGAND. « Improved H.264/AVC coding using texture analysis and synthesis ». In : *Image Processing (ICIP) International Conference on*. T. 3. IEEE. 2003, p. III–849.
- [464] Patrick NDJIKI-NYA, Bela MAKAI, Gabi BLATTERMANN, Aljoscha SMOLIC, Heiko SCHWARZ et Thomas WIEGAND. « Improved H.264/AVC coding using texture analysis and synthesis ». In : *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*. T. 3. IEEE. 2003, p. III–849.
- [465] Patrick NDJIKI-NYA et T WIEGAND. « Video coding using texture analysis and synthesis ». In : *Proc Picture Coding Symp. Saint-Malo, France*. 2003.

- [466] Atrde OLIVA, Antonio TORRALBA, Monica S CASTELHANO et John M HENDERSON. « Top-down control of visual attention in object detection ». In : *Image processing, 2003. icip 2003. proceedings. 2003 international conference on*. T. 1. IEEE. 2003, p. I-253.
- [467] EePing ONG, Weisi LIN, Zhongkang LU, Xiaokang YANG, Susu YAO, Xiao LIN et Fulvio MOSCHETTI. « Quality evaluation of MPEG-4 and H.26L coded video for mobile multimedia communications ». In : *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*. T. 1. IEEE. 2003, p. 473-476.
- [468] Iain E. RICHARDSON. *H.264 and MPEG-4 Video Compression : Video Coding for Next-generation Multimedia*. Wiley, 2003.
- [469] Chi-Wang TING, Lai-Man PO et Chun-Ho CHEUNG. « Center-biased frame selection algorithms for fast multi-frame motion estimation in H.264 ». In : *Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on*. T. 2. IEEE. 2003, p. 1258-1261.
- [470] S VALAEYS, Gloria MENEGAZ, Francesco ZILIANI et Julien REICHEL. « Modeling of 2D+1 texture movies for video coding ». In : *Image and Vision Computing 21.1 (2003)*, p. 49-59.
- [471] Khanh VU, Kien HUA, Wallapak TAVANAPONG et al. « Image retrieval based on regions of interest ». In : *Knowledge and Data Engineering, IEEE Transactions on 15.4 (2003)*, p. 1045-1049.
- [472] Yizhou WANG et Song-Chun ZHU. « Modeling textured motion : Particle, wave and sketch ». In : *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE. 2003, p. 213-220.
- [473] Thomas WIEGAND, Heiko SCHWARZ, Anthony JOCH, Faouzi KOSENTINI et Gary J SULLIVAN. « Rate-constrained coder control and comparison of video coding standards ». In : *Circuits and Systems for Video Technology, IEEE Transactions on 13.17 (2003)*, p. 688-703.
- [474] Chi-Wah WONG, Oscar C AU, Bojun MENG et Hong-Kwai LAM. « Perceptual rate control for low-delay video communications ». In : *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*. T. 3. IEEE. 2003, p. III-361.
- [475] Jianfeng XU, Ping YANG et Yun HE. « Modification of Fast Motion Estimation ». In : *JVT-J027. doc, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 10th Meeting, Waikoloa, HI, USA*. 2003.
- [476] Peng YIN, Alexis M TOURAPIS et Jill M BOYCE. « Fast mode decision and motion estimation for JVT/H.264 ». In : *ICIP*. IEEE. 2003, p. 853-856.

- 
- [6] Marc ANTONINI et Vincent RICORDEL. « Quantification ». In : *Compression et codage des images et des vidéos*. Sous la dir. de Michel BARLAUD et Claude LABIT. Traité IC2, Série traitement du signal et de l'image ISBN : 2-7462-0328-6. Hermes Science, jan. 2002. Chap. 2.
- [477] A CHAUVIN, J HERAULT, C MARENDAZ et C PEYRIN. « Natural scene perception : visual attractors and images processing ». In : *Progress in Neural Processing* (2002), p. 236–248.
- [478] Zhibo CHEN, Peng ZHOU, Yun HE et al. « Fast integer pel and fractional pel motion estimation for JVT ». In : *JVT-F017* (2002), p. 5–13.
- [479] Hye-Yeon CHEONG, AM TOURAPIS et P TOPIWALA. « Fast Motion Estimation within the JVT codec ». In : *5th Meeting of JVT of ISO/IEC MPEG and ITU-T VCEG, Geneva*. 2002.
- [480] Minh N DO et Martin VETTERLI. « Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance ». In : *Image Processing, IEEE Transactions on* 11.2 (2002), p. 146–158.
- [481] Erwan LE PENNEC. « Bandelettes et représentation géométrique des images ». Thèse de doct. École polytechnique, 2002.
- [482] Yan LI, Tianshu WANG et Heung-Yeung SHUM. « Motion texture : a two-level statistical model for character motion synthesis ». In : *ACM Transactions on Graphics (ToG)*. T. 21. 3. ACM. 2002, p. 465–472.
- [483] Xiuwen LIU et DeLiang WANG. « A spectral histogram model for texton modeling and texture discrimination ». In : *Vision Research* 42.23 (2002), p. 2617–2634.
- [484] Remi MEGRET et Daniel DEMENTHON. *A survey of spatio-temporal grouping techniques*. Rapp. tech. DTIC Document, 2002.
- [485] Timo OJALA, Matti PIETIKAINEN et Topi MAENPAA. « Multiresolution gray-scale and rotation invariant texture classification with local binary patterns ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7 (2002), p. 971–987.
- [486] Keun-hyeong PARK et HyunWook PARK. « Region-of-interest coding based on set partitioning in hierarchical trees ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 12.2 (2002), p. 106–113.
- [487] Derrick J PARKHURST. « Selective attention in natural vision : Using computational models to quantify stimulus-driven attentional allocation ». Thèse de doct. The Johns Hopkins University, 2002.
- [488] Daniel SCHARSTEIN et Richard SZELISKI. « A taxonomy and evaluation of dense two-frame stereo correspondence algorithms ». In : *International journal of computer vision, Springer* 47.1-3 (2002), p. 7–42.
- [489] John R SMITH, Ching-Yung LIN et Milind NAPHADE. « Video texture indexing using spatio-temporal wavelets ». In : *Image Processing. 2002. Proceedings. 2002 International Conference on*. T. 2. IEEE. 2002, p. II–437.



- [490] Jean-Luc STARCK, Emmanuel J CANDÈS et David L DONOHO. « The curvelet transform for image denoising ». In : *Image Processing, IEEE Transactions on* 11.6 (2002), p. 670–684.
- [491] SVT. *Overall-quality assessment when targeting Wide-XGA flat panel displays*. Rapp. tech. SVT corporate development technology, 2002.
- [492] Wenjun ZENG, Scott DALY et Shawmin LEI. « An overview of the visual optimization tools in JPEG 2000 ». In : *Signal Processing : Image Communication* 17.1 (2002), p. 85–104.
- [493] Ce ZHU, Xiao LIN et Lap-Pui CHAU. « Hexagon-based search pattern for fast block motion estimation ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 12.5 (2002), p. 349–355.
- [494] Gisle BJONTEGAARD. « Calculation of average PSNR differences between RD-curves ». In : *Doc. VCEG-M33 ITU-T Q6/16, Austin, TX, USA, 2-4 April 2001* (2001).
- [495] R. BOITARD, L. GUILLO et T. POIRIER. *Integration into the TMuC of an Intra Prediction based on a linear combination of Template Matching predictors*. Joint Collaborative Team on Video Coding, JCTVC-D193. 2001.
- [496] Laurent ITTI et Christof KOCH. « Computational modelling of visual attention ». In : *Nature reviews neuroscience* 2.3 (2001), p. 194–203.
- [497] Stanley A KLEIN. « Measuring, estimating, and understanding the psychometric function : A commentary ». In : *Perception & psychophysics* 63.8 (2001), p. 1421–1455.
- [498] Patrick LE CALLET. « Critères objectifs avec référence de qualité visuelle des images couleur ». Thèse de doct. Université de Nantes, Ecole polytechnique de l’uni- versité de Nantes, 2001.
- [499] Marjorie R LEEK. « Adaptive procedures in psychophysical research ». In : *Perception & psychophysics* 63.8 (2001), p. 1279–1292.
- [500] Christopher PACK, Stephen GROSSBERG et Ennio MINGOLLA. « A neural model of smooth pursuit control and motion perception by cortical area MST ». In : *Journal of Cognitive Neuroscience* 13.1 (2001), p. 102–120.
- [501] Sundeep RANGAN et Vivek K GOYAL. « Recursive consistent estimation with bounded noise ». In : *Information Theory, IEEE Transactions on* 47.1 (2001), p. 457–464.
- [25] Vincent RICORDEL, Sébastien PARIS et Jan OPDERBECKE. « Trajectory Estimation for Ultrashort Baseline Acoustic Positioning Systems ». In : *IEEE Aerospace Conference*. Big Sky, Montana, USA. Mar. 2001.
- [502] Stefano SOATTO, Gianfranco DORETTO et Ying Nian WU. « Dynamic textures ». In : *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. T. 2. IEEE. 2001, p. 439–446.

- 
- [503] Yao WANG, Michael T ORCHARD, Vinay VAISHAMPAYAN et Amy R REIBMAN. « Multiple description coding using pairwise correlating transforms ». In : *Image Processing, IEEE Transactions on* 10.3 (2001), p. 351–366.
- [504] Felix A WICHMANN et N Jeremy HILL. « The psychometric function : I. Fitting, sampling, and goodness of fit ». In : *Perception & psychophysics* 63.8 (2001), p. 1293–1313.
- [505] Hector YEE, Sumanita PATTANAIK et Donald P GREENBERG. « Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments ». In : *ACM Transactions on Graphics (TOG)* 20.1 (2001), p. 39–65.
- [506] Scott J DALY, Wenjun ZENG, Jin Li et Shawmin LEI. « Visual masking in wavelet compression for JPEG-2000 ». In : *Electronic Imaging*. International Society for Optics et Photonics. 2000, p. 66–80.
- [507] Russell L DE VALOIS, Nicolas P COTTARIS, Luke E MAHON, Sylvia D ELFAR et J Anthony WILSON. « Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity ». In : *Vision research* 40.27 (2000), p. 3685–3702.
- [508] Minh N DO et Martin VETTERLI. « Texture similarity measurement using Kullback-Leibler distance on wavelet subbands ». In : *Image Processing, 2000. Proceedings. 2000 International Conference on*. T. 3. IEEE. 2000, p. 730–733.
- [509] Peter G ENGELDRUM. *Psychometric scaling : a toolkit for imaging systems development*. Imcotek press, 2000.
- [510] Stephen P ETZ et Jiebo LUO. « Ground truth for training and evaluation of automatic main subject detection ». In : *Electronic Imaging*. ISOP. 2000, p. 434–442.
- [511] Laurent ITTI et Christof KOCH. « A saliency-based search mechanism for overt and covert shifts of visual attention ». In : *Vision research* 40.10 (2000), p. 1489–1506.
- [512] Laurent ITTI, Christof KOCH et Jochen BRAUN. « Revisiting spatial vision : Toward a unifying model ». In : *JOSA A* 17.11 (2000), p. 1899–1917.
- [513] Jiebo LUO et Amit SINGHAL. « On measuring low-level saliency in photographic images ». In : *CVPR*. IEEE. 2000.
- [514] Atsuto MAKI, Peter NORDLUND et Jan-Olof EKLUNDH. « Attentional scene segmentation : integrating depth and motion ». In : *Computer Vision and Image Understanding*. Sous la dir. d’ELSEVIER. T. 78. 3. 2000, p. 351–373.
- [43] Jean MICHELIS, Vincent RICORDEL, William PUECH, Guy PASSAIL et Michel DUMAS. « Intégration d’applet JAVA dans un réseau d’images médicales : aide au télédiagnostic ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Poitiers, France. Oct. 2000.

- [515] Alexander E MOHR, Eve RISKIN, Richard E LADNER et al. « Unequal loss protection : graceful degradation of image quality over packet erasure channels through forward error correction ». In : *Selected Areas in Communications, IEEE Journal on* 18.6 (2000), p. 819–828.
- [516] Javier PORTILLA et Eero P SIMONCELLI. « A parametric texture model based on joint statistics of complex wavelet coefficients ». In : *International Journal of Computer Vision* 40.1 (2000), p. 49–70.
- [517] Javier PORTILLA et Eero P SIMONCELLI. « A parametric texture model based on joint statistics of complex wavelet coefficients ». In : *International Journal of Computer Vision* 40.1 (2000), p. 49–70.
- [518] Claudio M PRIVITERA et Lawrence W STARK. « Algorithms for defining visual regions-of-interest : Comparison with eye fixations ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.9 (2000), p. 970–982.
- [26] William PUECH, Guy PASSAIL et Vincent RICORDEL. « Analysis and Optimisation of 3D Reconstruction Method of the Aorta from a Tomographic Image Sequence ». In : *European Signal Processing Conference (EUSIPCO)*. Tampere, Finlande. Sept. 2000.
- [519] David TAUBMAN. « High performance scalable image compression with EBCOT ». In : *Image Processing, IEEE transactions on* 9.7 (2000), p. 1158–1170.
- [520] Alexis M TOURAPIS, Oscar CL AU et Ming L LIOU. « Predictive motion vector field adaptive search technique (PMVFAST) : enhancing block-based motion estimation ». In : *Photonics West 2001-Electronic Imaging*. International Society for Optics et Photonics. 2000, p. 883–892.
- [521] Yiwei WANG, John E DOHERTY et Robert E VAN DYCK. « Moving object tracking in video ». In : *Applied Imagery Pattern Recognition Workshop, 2000. Proceedings. 29th*. IEEE. 2000, p. 95–101.
- [522] Jeremy WOLFE. « Visual attention ». In : *Seeing* 2.335–386 (2000).
- [523] Wenjun ZENG, Scott DALY et Shawmin LEI. « Point-wise extended visual masking for JPEG-2000 image compression ». In : *Image Processing, 2000. Proceedings. 2000 International Conference on*. T. 1. IEEE. 2000, p. 657–660.
- [524] Peter GJ BARTEN. *Contrast sensitivity of the human eye and its effects on image quality*. T. 21. Spie optical engineering press Bellingham, WA, 1999.
- [525] Emmanuel J CANDÈS et David L DONOHO. « Ridgelets : A key to higher-dimensional intermittency? » In : *Philosophical Transactions of the Royal Society of London A : Mathematical, Physical and Engineering Sciences* 357.1760 (1999), p. 2495–2509.
- [526] T. GRANDPIERRE, C. LAVARENNE et Y. SOREL. « Optimized Rapid Prototyping For Real-Time Embedded Heterogeneous multiprocessors ». In : *Proceedings of 7th International Workshop on Hardware/Software Co-Design (CODES)*. ACM. Rome, Italy, mai 1999, p. 74–78.

- [527] Stephen GROSSBERG, Ennio MINGOLLA et Christopher PACK. « A neural model of motion processing and visual navigation by cortical area MST ». In : *Cerebral Cortex* 9.8 (1999), p. 878–895.
- [528] John M HENDERSON, Phillip A WEEKS JR et Andrew HOLLINGWORTH. « The effects of semantic consistency on eye movements during complex scene viewing ». In : *Journal of experimental psychology : Human perception and performance* 25.1 (1999), p. 210.
- [529] Raymond M KLEIN et W Joseph MACINNES. « Inhibition of return is a foraging facilitator in visual search ». In : *Psychological science* 10.4 (1999), p. 346–352.
- [530] Austin Y LAN, Anthony G NGUYEN et Jenq-Neng HWANG. « Scene-context-dependent reference-frame placement for MPEG video coding ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 9.3 (1999), p. 478–489.
- [44] Guy PASSAIL, Vincent RICORDEL, Alexander KOLESNIKOV et William PUECH. « Mise en œuvre d’un réseau pour la communication d’images médicales ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Nice-Sophia-Antipolis, France. Juin 1999.
- [45] William PUECH, Guy PASSAIL, Steeve NICOLAY et Vincent RICORDEL. « Analyse et amélioration de méthodes de reconstruction 3D de l’aorte à partir d’une séquence d’images tomographiques ». In : *Colloque GRETSI sur le traitement du signal et des images (GRETSI)*. Vannes, France. Sept. 1999.
- [531] Rohit PURI et Kannan RAMCHANDRAN. « Multiple description source coding using forward error correction ». In : *Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on*. T. 1. IEEE. 1999, p. 342–346.
- [532] ITU-T RECOMMENDATION. « Subjective video quality assessment methods for multimedia applications ». In : *International telecommunication union* (1999).
- [533] Pamela REINAGEL et Anthony M ZADOR. « Natural scene statistics at the center of gaze ». In : *Network : Computation in Neural Systems* 10.4 (1999), p. 1–10.
- [534] Jarek ROSSIGNAC. « Edgebreaker : Connectivity compression for triangle meshes ». In : *IEEE transactions on visualization and computer graphics* 5.1 (1999), p. 47–61.
- [535] Stefan WINKLER. « Perceptual distortion metric for digital color video ». In : *Electronic Imaging’99*. International Society for Optics et Photonics. 1999, p. 175–184.
- [536] Laurent BEDAT. « Aspects psychovisuels de la perception des couleurs. Application au codage d’images couleurs fixes avec compression de l’information ». Thèse de doct. Université de Nantes, IRESTE, 1998.
- [537] Emmanuel Jean CANDES. « Ridgelets : theory and applications ». Thèse de doct. Stanford University, 1998.

- [538] Scott J DALY. « Engineering observations from spatiovelocity and spatiotemporal visual models ». In : *Photonics West'98 Electronic Imaging*. International Society for Optics et Photonics. 1998, p. 180–191.
- [539] Ingrid DAUBECHIES et Wim SWELDENS. « Factoring wavelet transforms into lifting steps ». In : *Journal of Fourier analysis and applications* 4.3 (1998), p. 247–269.
- [540] RE FREDERICKSEN et RF HESS. « Estimating multiple temporal mechanisms in human vision ». In : *Vision Research* 38.7 (1998), p. 1023–1040.
- [541] Vivek K GOYAL, Jelena KOVACEVIC et Martin VETTERLI. « Multiple description transform coding : Robustness to erasures using tight frame expansions ». In : *IEEE International Symposium on Information Theory*. IEEE. 1998, p. 408.
- [542] Laurent ITTI, Christof KOCH et Ernst NIEBUR. « A model of saliency-based visual attention for rapid scene analysis ». In : *IEEE Transactions on Pattern Analysis & Machine Intelligence* 11 (1998), p. 1254–1259.
- [543] Wilfried M OSBERGER, Anthony J MAEDER et Neil W BERGMANN. « Perceptually based quantization technique for mpeg encoding ». In : *Photonics West'98 Electronic Imaging*. International Society for Optics et Photonics. 1998, p. 148–159.
- [544] Wilfried OSBERGER et Anthony J MAEDER. « Automatic identification of perceptually important regions in an image ». In : *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*. T. 1. IEEE. 1998, p. 701–704.
- [27] Vincent RICORDEL et Moncef GABBOUJ. « Adaptive Tree-Structure Lattice Vector Quantization for Video Coding ». In : *European Signal Processing Conference (EUSIPCO)*. Rhodes, Grèce. Sept. 1998.
- [46] Vincent RICORDEL, Moncef GABBOUJ et Claude LABIT. « Quantification vectorielle algébrique et arborescente : méthode d'actualisation du dictionnaire. Application au codage vidéo ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Lannion, France. Juin 1998.
- [545] Eero P SIMONCELLI et David J HEEGER. « A model of neuronal responses in visual area MT ». In : *Vision research* 38.5 (1998), p. 743–761.
- [546] Gary J SULLIVAN et Thomas WIEGAND. « Rate-distortion optimization for video compression ». In : *Signal Processing Magazine, IEEE* 15.6 (1998), p. 74–90.
- [547] Gabriel TAUBIN et Jarek ROSSIGNAC. « Geometric compression through topological surgery ». In : *ACM Transactions on Graphics (TOG)* 17.2 (1998), p. 84–115.
- [548] Costa TOUMA. « Triangle mesh compression ». In : *Graphics Interface 98*. 1998.
- [549] TUCERYAN, MIHRAN AND JAIN, ANIL K. « Texture analysis ». In : *The handbook of pattern recognition and computer vision* 2 (1998), p. 207–248.

- 
- [550] ITU International Telecommunication UNION. *Subjective assessment methods for image quality in high-definition television*. Recommendation ITU-R BT.710-4. 1998.
- [551] Andrew B WATSON. « Toward a perceptual video-quality metric ». In : *Photonics West'98 Electronic Imaging*. International Society for Optics et Photonics. 1998, p. 139–147.
- [552] Henrique S MALVAR. « Lapped biorthogonal transforms for transform coding with reduced blocking and ringing artifacts ». In : *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*. T. 3. IEEE. 1997, p. 2421–2424.
- [553] SK MANNAN, KH RUDDOCK et DS WOODING. « Fixation sequences made during visual examination of briefly presented 2D images ». In : *Spatial Vision* 11.2 (1997), p. 157–178.
- [554] Willfried OSBERGER, Sean HAMMOND et Neil BERGMANN. « An MPEG encoder incorporating perceptually based quantisation ». In : *TENCON'97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications., Proceedings of IEEE*. T. 2. IEEE. 1997, p. 731–734.
- [555] Yao WANG, Michael T ORCHARD et Amy R REIBMAN. « Multiple description image coding for noisy channels by pairing transform coefficients ». In : *Multimedia Signal Processing, 1997., IEEE First Workshop on*. IEEE. 1997, p. 419–424.
- [556] MJ HAWKEN, RM SHAPLEY et DH GROSOF. « Temporal-frequency selectivity in monkey visual cortex ». In : *Visual neuroscience* 13.3 (1996), p. 477–492.
- [557] Christian J Van Den Branden LAMBRECHT. « Perceptual models and architectures for video coding applications ». Thèse de doct. Citeseer, 1996.
- [558] Lurng-Kuo LIU et Ephraim FEIG. « A block-based gradient descent search algorithm for block motion estimation in video coding ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 6.4 (1996), p. 419–422.
- [559] Frank H MAHNKE. *Color, environment, and human response : an interdisciplinary understanding of color and its use as a beneficial element in the design of the architectural environment*. John Wiley & Sons, 1996.
- [560] Atsuto MAKI, Peter NORDLUND et Jan-Olof EKLUNDH. « A computational model of depth-based attention ». In : *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*. T. 4. IEEE. 1996, p. 734–739.
- [561] Bangalore S MANJUNATH et Wei-Ying MA. « Texture features for browsing and retrieval of image data ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 18.8 (1996), p. 837–842.
- [562] Xavier MARICHAL, Thierry DELMOT, Christophe DE VLEESCHOUWER, Vincent WARSCOTTE, Benoit MACQ et al. « Automatic detection of interest areas of an image or of a sequence of images ». In : *ICIP*. IEEE. 1996, p. 371–374.

- [563] Jonathan A MARSHALL, Christina A BURBECK, Dan ARIELY, Jannick P ROLLAND et Kevin E MARTIN. « Occlusion edge blur : a cue to relative visual depth ». In : *JOSA A* 13.4 (1996), p. 681–688.
- [564] George MATHER. « Image blur as a pictorial depth cue ». In : *Proceedings of the Royal Society of London B : Biological Sciences* 263.1367 (1996), p. 169–172.
- [565] Lai-Man PO et Wing-Chung MA. « A novel four-step search algorithm for fast block motion estimation ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 6.3 (1996), p. 313–317.
- [47] Vincent RICORDEL et Claude LABIT. « Quantification vectorielle algébrique et arborescente ». In : *Compression et représentation des signaux audiovisuels (CORESA)*. Grenoble, France. Fév. 1996.
- [28] Vincent RICORDEL et Claude LABIT. « Tree-Structured Lattice Vector Quantization ». In : *European Signal Processing Conference (EUSIPCO)*. Trieste, Italie. Sept. 1996.
- [566] Hakim SÉNANE. « Représentation d’images en sous-bandes visuelles. Application au codage d’images de télévision sans défaut visuel ». Thèse de doct. Université de Nantes, IRESTE, 1996.
- [567] S YANTIS et J JONIDAS. « Attentional capture by abrupt onsets and selective attention : evidence from visual search ». In : *J. Experimental Psychology : Human Perception Performance* 20 (1996), p. 1505–1513.
- [568] Jiyong ZHAO, Yoshihisa SHIMAZU, Koji OHTA, Rina HAYASAKA et Yuki MATSUHITA. « An outstandingness oriented image segmentation and its application ». In : *Signal Processing and Its Applications, 1996. ISSPA 96., Fourth International Symposium on*. T. 1. IEEE. 1996, p. 45–48.
- [569] Chun-Hsien CHOU et Yun-Chin LI. « A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 5.6 (1995), p. 467–476.
- [570] Charles KERVRANN et Fabrice HEITZ. « A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics ». In : *Image Processing, IEEE Transactions on* 4.6 (1995), p. 856–862.
- [571] Anthony J MAEDER. « Importance maps for adaptive information reduction in visual scenes ». In : *Intelligent Information Systems, 1995. ANZIIS-95. Proceedings of the Third Australian and New Zealand Conference on*. IEEE. 1995, p. 24–29.
- [572] David F MCALLISTER, Susan MCKAY et al. « Stereo computer graphics and other true 3D technologies ». In : *Computers in Physics* 9.1 (1995), p. 56.
- [573] Eric NGUYEN-PHUC. « Compression sélective et focalisation visuelle : application au codage hybride de séquences d’images ». Thèse de doct. Université Rennes 1, 1995.

- [574] Jean-Marc ODOBEZ et Patrick BOUTHEMY. « Robust multiresolution estimation of parametric motion models ». In : *Journal of visual communication and image representation*. T. 6. 4. Elsevier, 1995, p. 348–365.
- [48] Vincent RICORDEL et Claude LABIT. « Quantification vectorielle par emboitage de réseaux réguliers de points ». In : *Colloque GRETSI sur le traitement du signal et des images (GRETSI)*. Juans-les-Pins, France. Sept. 1995.
- [29] Vincent RICORDEL et Claude LABIT. « Vector Quantization by Hierarchical Packing of Embedded Truncated Lattices ». In : *Visual Communications and Image Processing (VCIP)*. Taiwan, Chine. Mai 1995.
- [30] Vincent RICORDEL et Claude LABIT. « Vector Quantization by packing of embedded truncated lattices ». In : *IEEE International Conference on Image Processing (ICIP)*. Washington DC, USA. Oct. 1995.
- [575] Thomas SIKORA et Bela MAKAI. « Shape-adaptive DCT for generic coding of video ». In : *IEEE Transactions on Circuits and Systems for Video Technology* 5.1 (1995), p. 59–62.
- [576] Wa James TAM, Lew B STELMACH, Limin WANG, Daniel LAUZON et Peter GRAY. « Visual masking at video scene cuts ». In : *IS&T/SPIE's Symposium on Electronic Imaging : Science & Technology*. International Society for Optics et Photonics. 1995, p. 111–119.
- [577] Bernhard TREUTWEIN. « Adaptive psychophysical procedures ». In : *Vision research* 35.17 (1995), p. 2503–2522.
- [578] John K TSOTSOS, Scan M CULHANE, Winky Yan Kei WAI, Yuzhong LAI, Neal DAVIS et Fernando NUFLO. « Modeling visual attention via selective tuning ». In : *Artificial intelligence* 78.1 (1995), p. 507–545.
- [579] John M FOLEY et Geoffrey M BOYNTON. « New model of human luminance pattern vision mechanisms : analysis of the effects of pattern orientation, spatial phase, and temporal frequency ». In : *Computational Vision Based on Neurobiology*. International Society for Optics et Photonics. 1994, p. 32–42.
- [580] Jungwoo LEE et Bradley W DICKINSON. « Temporally adaptive motion interpolation exploiting temporal masking in visual perception ». In : *Image Processing, IEEE Transactions on* 3.5 (1994), p. 513–526.
- [581] Andrew B WATSON. « Perceptual optimization of DCT color quantization matrices ». In : *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*. T. 1. IEEE. 1994, p. 100–104.
- [582] AJ AHUMADA et Cynthia H NULL. « Image quality : A multidimensional problem ». In : *Digital images and human vision* (1993), p. 141–148.
- [583] KW CHUN, KW LIM, HD CHO et JB RA. « An adaptive perceptual quantization algorithm for video coding ». In : *Consumer Electronics, IEEE Transactions on* 39.3 (1993), p. 555–558.



- [584] Russell L DE VALOIS et Karen K DE VALOIS. « A multi-stage color model ». In : *Vision research* 33.8 (1993), p. 1053–1065.
- [585] David J HEEGER. « Modeling simple-cell direction selectivity with normalized, half-squared, linear operators ». In : *Journal of Neurophysiology* 70.5 (1993), p. 1885–1898.
- [586] Nikil JAYANT, James JOHNSTON et Robert SAFRANEK. « Signal compression based on models of human perception ». In : *Proceedings of the IEEE* 81.10 (1993), p. 1385–1422.
- [587] Ruggero MILANESE. « Detecting salient regions in an image : from biological evidence to computer implementation ». Thèse de doct. University of Geneva, 1993.
- [588] Eli PELI, Lawrence E AREND, George M YOUNG et Robert B GOLDSTEIN. « Contrast sensitivity to patch stimuli : Effects of spatial bandwidth and temporal presentation ». In : *Spatial Vision* 7.1 (1993), p. 1–14.
- [589] Vinay Anant VAISHAMPAYAN. « Design of multiple description scalar quantizers ». In : *Information Theory, IEEE Transactions on* 39.3 (1993), p. 821–834.
- [590] Andrew B WATSON. « DCTune : A technique for visual optimization of DCT quantization matrices for individual images ». In : *Sid International Symposium Digest of Technical Papers*. T. 24. SOCIETY FOR INFORMATION DISPLAY. 1993, p. 946–946.
- [591] Arthur A WEBSTER, Coleen T JONES, Margaret H PINSON, Stephen D VORAN et Stephen WOLF. « Objective video quality assessment system based on human perception ». In : *IS&T/SPIE's Symposium on Electronic Imaging : Science and Technology*. International Society for Optics and Photonics. 1993, p. 15–26.
- [592] Scott J DALY. « Visible differences predictor : an algorithm for the assessment of image fidelity ». In : *SPIE/IS&T 1992 Symposium on Electronic Imaging : Science and Technology*. International Society for Optics et Photonics. 1992, p. 2–15.
- [593] Melvyn A GOODALE et A David MILNER. « Separate visual pathways for perception and action ». In : *Trends in neurosciences* 15.1 (1992), p. 20–25.
- [594] Ruggero MILANESE, Jean-Marc BOST et Thierry PUN. « A bottom-up attention system for active vision ». In : *Proceedings of the 10th European conference on Artificial intelligence*. John Wiley & Sons, Inc. 1992, p. 808–810.
- [595] Randal C NELSON et Ramprasad POLANA. « Qualitative recognition of motion using temporal texture ». In : *CVGIP : Image understanding* 56.1 (1992), p. 78–89.
- [596] Eero P SIMONCELLI, William T FREEMAN, Edward H ADELSON et David J HEEGER. « Shiftable multiscale transforms ». In : *Information Theory, IEEE Transactions on* 38.2 (1992), p. 587–607.

- 
- [597] Hiroshi WATANABE et S SINBGHAL. « Bit allocation and rate control based on human visual sensitivity for interframe coders ». In : *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*. T. 3. IEEE. 1992, p. 521–524.
- [598] Dana H BALLARD. « Animate vision ». In : *Artificial intelligence* 48.1 (1991), p. 57–86.
- [599] Eve A RISKIN. « Optimal bit allocation via the generalized BFOS algorithm ». In : *IEEE Transactions on information Theory* 37.2 (1991), p. 400–402.
- [600] Eve A RISKIN et Robert M GRAY. « A greedy tree growing algorithm for the design of variable rate vector quantizers (image compression) ». In : *IEEE Transactions on Signal Processing* 39.11 (1991), p. 2500–2507.
- [601] Luc VINCENT et Pierre SOILLE. « Watersheds in digital spaces : an efficient algorithm based on immersion simulations ». In : *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6 (1991), p. 583–598.
- [602] Jochen BRAUN et Dov SAGI. « Vision outside the focus of attention ». In : *Perception & Psychophysics* 48.1 (1990), p. 45–58.
- [603] Patrick FLANAGAN, Patrick CAVANAGH et Olga Eizner FAVREAU. « Independent orientation-selective mechanisms for the cardinal directions of colour space ». In : *Vision research* 30.5 (1990), p. 769–778.
- [604] Dong-Chen HE et Li WANG. « Texture unit, texture spectrum, and texture analysis ». In : *Geoscience and Remote Sensing, IEEE Transactions on* 28.4 (1990), p. 509–512.
- [605] Patrick LALANDE. « Detection du mouvement dans une sequence d’images selon une approche markovienne ; application a la robotique sous-marine ». Thèse de doct. Université Rennes 1, 1990.
- [606] Michael A WEBSTER, Eugene SWITKES et Karen K DE VALOIS. « Orientation and spatial-frequency discrimination for luminance and chromatic gratings ». In : *JOSA A* 7.6 (1990), p. 1034–1049.
- [607] Moses AMADASUN et Robert KING. « Textural features corresponding to textural properties ». In : *Systems, Man and Cybernetics, IEEE Transactions on* 19.5 (1989), p. 1264–1274.
- [608] Peter J BOWLER. *Evolution : the history of an idea*. Univ of California Press, 1989.
- [609] Philip A. CHOU, Tom LOOKABAUGH et Robert M. GRAY. « Optimal pruning with applications to tree-structured source coding and modeling ». In : *IEEE transactions on information theory* 35.2 (1989), p. 299–315.
- [610] Bernd GIROD. « The information theoretical significance of spatial and temporal masking in video signals ». In : *OE/LASE’89, 15-20 Jan., Los Angeles. CA*. International Society for Optics et Photonics. 1989, p. 178–189.

- [611] Jeremy M WOLFE, Kyle R CAVE et Susan L FRANZEL. « Guided search : an alternative to the feature integration model for visual search ». In : *Journal of Experimental Psychology : Human perception and performance* 15.3 (1989), p. 419.
- [612] Paul B CHOU et Rajeev RAMAN. *On Relaxation Algorithms Based on Markov Random Fields*. Rapp. tech. DTIC Document, 1987.
- [613] Christof KOCH et Shimon ULLMAN. « Shifts in selective visual attention : towards the underlying neural circuitry ». In : *Matters of intelligence*. Springer, 1987, p. 115–141.
- [614] Alex Paul PENTLAND. « A new sense for depth of field ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 4 (1987), p. 523–531.
- [615] Andrew B WATSON. « The cortex transform : rapid computation of simulated neural images ». In : *Computer vision, graphics, and image processing* 39.3 (1987), p. 311–327.
- [616] Mark R TURNER. « Texture discrimination by Gabor functions ». In : *Biological cybernetics* 55.2-3 (1986), p. 71–82.
- [617] KH FOSTER, James P GASKA, M NAGLER et DA POLLEN. « Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. » In : *The Journal of physiology* 365.1 (1985), p. 331–363.
- [618] Stuart GEMAN et Donald GEMAN. « Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 6 (1984), p. 721–741.
- [619] Michael I POSNER et Yoav COHEN. « Components of visual orienting ». In : *Attention and performance X : Control of language processes* 32 (1984), p. 531–556.
- [620] Jacob BECK. « Textural segmentation, second-order statistics, and textural elements ». In : *Biological Cybernetics* 48.2 (1983), p. 125–130.
- [621] Peter J BURT et Edward H ADELSON. « The Laplacian pyramid as a compact image code ». In : *Communications, IEEE Transactions on* 31.4 (1983), p. 532–540.
- [622] Scott KIRKPATRICK, C Daniel GELATT, Mario P VECCHI et al. « Optimization by simulated annealing ». In : *science* 4598 (1983), p. 671–680.
- [623] Daniel A POLLEN et Steven F RONNER. « Visual cortical neurons as localized spatial frequency filters ». In : *Systems, Man and Cybernetics, IEEE Transactions on* 5 (1983), p. 907–916.
- [624] John KRAUSKOPF, David R WILLIAMS et David W HEELEY. « Cardinal directions of color space ». In : *Vision research* 22.9 (1982), p. 1123–1131.
- [625] Berthold K HORN et Brian G SCHUNCK. « Determining optical flow ». In : *1981 Technical symposium east*. T. 319–331. International Society for Optics et Photonics. 1981.

- [626] Jaswant R JAIN et Anil K JAIN. « Displacement measurement and its application in interframe image coding ». In : *Communications, IEEE Transactions on* 29.12 (1981), p. 1799–1808.
- [627] Bela JULESZ. « Textons, the elements of texture perception, and their interactions ». In : *Nature* 290.5802 (1981), p. 91–97.
- [628] T KOGA. « Motion-compensated interframe coding for video conferencing ». In : *Proc. NTC 81*. 1981, p. C9–6.
- [629] P HAUBNER, H-W BODMANN et AM MARSDEN. « A unified relationship between brightness and luminance ». In : *Siemens Forschungs und Entwicklungsberichte* 9 (1980), p. 315–318.
- [630] Michael I POSNER. « Orienting of attention ». In : *Quarterly journal of experimental psychology* 32.1 (1980), p. 3–25.
- [631] DO REUDINK. « The channel splitting problem with interpolative coders ». In : *Rapport technique TM80* 134.1 (1980).
- [632] Anne M TREISMAN et Garry GELADE. « A feature-integration theory of attention ». In : *Cognitive psychology* 12.1 (1980), p. 97–136.
- [633] O BRADDICK, FW CAMPBELL et J ATKINSON. « Channels in vision : Basic aspects ». In : *Perception*. Springer, 1978, p. 3–38.
- [634] Bela JULESZ, EN GILBERT et Jonathan D VICTOR. « Visual discrimination of textures with identical third-order statistics ». In : *Biological Cybernetics* 31.3 (1978), p. 137–140.
- [635] Hideyuki TAMURA, Shunji MORI et Takashi YAMAWAKI. « Textural features corresponding to visual perception ». In : *Systems, Man and Cybernetics, IEEE Transactions on* 8.6 (1978), p. 460–473.
- [636] Olivier Dominique FAUGERAS. « Digital color image processing and psychophysics within the framework of a human visual model ». Thèse de doct. University of Utah, 1976.
- [637] Julian BESAG. « Spatial interaction and the statistical analysis of lattice systems ». In : *Journal of the Royal Statistical Society. Series B (Methodological)* (1974), p. 192–236.
- [638] James L MANNOS et David J SAKRISON. « The effects of a visual fidelity criterion of the encoding of images ». In : *Information Theory, IEEE Transactions on* 20.4 (1974), p. 525–536.
- [639] Bela JULÉSZ, EN GILBERT, LA SHEPP et HL FRISCH. « Inability of humans to discriminate between visual textures that agree in second-order statistics-revisited ». In : *Perception* 2.4 (1973), p. 391–405.
- [640] Johannes ITTEN et Faber BIRREN. *The elements of color*. T. 4. John Wiley & Sons, 1970.

- 
- [641] David H HUBEL et Torsten N WIESEL. « Receptive fields and functional architecture of monkey striate cortex ». In : *The Journal of physiology* 195.1 (1968), p. 215–243.
- [642] MiM TAYLOR et C Douglas CREELMAN. « PEST : Efficient estimates on probability functions ». In : *The Journal of the Acoustical Society of America* 41.4A (1967), p. 782–787.
- [643] Alfred L YARBUS. *Eye movements during perception of complex objects*. Springer, 1967.
- [644] Albert J SEYLER et Zigmantas L BUDRIKIS. « Detail perception after scene changes in television image presentations ». In : *Information Theory, IEEE Transactions on* 11.1 (1965), p. 31–43.
- [645] Bela JULESZ. « Visual pattern discrimination ». In : *IRE transactions on Information Theory* 8.2 (1962), p. 84–92.
- [646] AJ SEYLER et ZL BUDRIKIS. « Measurements of temporal adaptation to spatial detail vision ». In : *Nature* 184 (1959), p. 1215–1217.
- [647] H DE LANGE DZN. « Research into the dynamic nature of the human fovea, Ūi cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light ». In : *Josa* 48.11 (1958), p. 777–783.
- [648] Sidney M NEWHALL, Dorothy NICKERSON et Deane B JUDD. « Final report of the OSA subcommittee on the spacing of the munsell colors ». In : *JOSA* 33.7 (1943), p. 385–411.