

Projet TEMICS

Traitement, Modélisation d'Images et CommunicationS

Rennes

THÈME 3B



*R*apport
d'Activité

2001

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	4
3	Fondements scientifiques	5
3.1	Modélisation 3D par géométrie projective	5
3.2	Estimation de mouvement et segmentation	6
3.3	Approximation par maillages hiérarchiques et éléments finis	7
3.4	Théorie débit-distorsion et formalisme MDL	9
3.5	Représentation temps-fréquence et par bases de fonctions redondantes	10
3.6	Tatouage d'images	12
4	Domaines d'applications	13
4.1	Panorama	13
4.2	Compression d'images fixes et animées avec fonctionnalités avancées	14
4.3	Communication multimédia	14
4.4	Protection, authentification et enrichissement de contenus images et vidéo	15
5	Logiciels	16
5.1	Plate-forme de communication vidéo	16
5.2	Codecs vidéo et interface réseau	16
6	Résultats nouveaux	17
6.1	Analyse et modélisation de séquences d'images	17
6.1.1	Modélisation 3D par analyse de séquences vidéo monoculaires	17
6.1.2	Segmentation d'objet vidéo pour le codage	18
6.1.3	Représentation par maillages et mosaïques généralisées d'objets vidéo	18
6.1.4	Analyse des conditions d'illumination dans une séquence d'images	20
6.1.5	Structuration par analyse des trajectoires des objets vidéo	20
6.2	Codage et codage conjoint source-canal	22
6.2.1	Codage orienté modèles 3D	23
6.2.2	Codage scalable à grain fin	24
6.2.3	Contrôle de congestion et régulation de débit	25
6.2.4	Codage conjoint source-canal	25
6.2.5	Décodage conjoint source-canal	27
6.3	Tatouage d'images	27
6.4	Résultats relatifs à l'action TexMex en cours de montage	28
7	Contrats industriels (nationaux, européens et internationaux)	29
7.1	Contrats industriels bi-latéraux	29
7.1.1	CTI-France Télécom R&D	29
7.1.2	Convention Celar	29
7.1.3	Convention Thomson-TCO	30

7.1.4	Convention Thomson Multimédia	30
7.2	Contrats nationaux	30
7.2.1	RNRT-V2net	30
7.2.2	RNRT-Osiam	31
7.2.3	RNRT-Cohrainte	31
7.2.4	RNRT-Cosocati	32
7.2.5	RNRT-Visi	32
7.3	Contrats européens	33
7.3.1	IST-Openise	33
7.3.2	IST-Song	33
8	Actions régionales, nationales et internationales	34
8.1	Réseaux et groupes de travail internationaux bilatéraux	34
8.2	Actions nationales	34
8.2.1	ACI Jeunes chercheurs	34
8.2.2	Action CNRS JemSTIC	35
9	Diffusion de résultats	35
9.1	Animation de la communauté scientifique	35
9.2	Enseignement	36
10	Bibliographie	36

1 Composition de l'équipe

Responsable scientifique

Christine Guillemot [DR Inria]

Assistante de projet

Edith Blin-Guyot [TR Inria]

Personnel Inria

Claude Labit [DR, 1/5 temps depuis mars 1999]

Henri Nicolas [CR]

Stéphane Pateux [CR, détachement du corps des Télécoms]

Pierre Siohan [DR, détachement de France Télécom]

Personnel CNRS

Laurent Amsaleg [CR]

Thierry Fauconnier [IR, (Atelier)]

Personnel Université

Luce Morin [maître de conférences, université Rennes 1]

Collaborateur Extérieur

Francisco Garcia-Ugalde [PAST]

Ingénieurs experts

Bilal El Ali [IE Inria]

Gabriele Dawini [IE Inria, à partir du 1er décembre 2001]

Chercheurs post-doctorants

Zhikui Chen [Post-doctorat jusqu'au 31 Mai 2000]

Teddy Furon [Post-doctorat UCL à partir du 1er Octobre 2001]

Slavinca Marinkovic [Post-doctorat à partir du 1er Novembre 2001]

Gagan Rath [Post-doctorat]

Ingénieurs associés

Philippe Coval [à partir du 1er octobre 2000]

Jeanne Guillou [CNRT-TIM, jusqu'au 31 octobre 2001]

Brendan Catherine [à partir du 1er octobre 2001]

Chercheurs doctorants

Sid-Ahmed Berrani [Bourse CIFRE Thomson Multimedia, 50% TEMICS - 50% VISTA]

Nathalie Camas [Bourse CIFRE France Télécom, à partir du 6 Novembre 2001]

Dubhe Chavira-Martinez [Bourse CONACYT]

Marc Chaumont [Bourse MENRT, à partir du 1er octobre 2000]

Franck Denoual [Bourse CIFRE, jusqu'au 31 Mai 2001]

Franck Galpin [Bourse INRIA-RÉGION, jusqu'au 31 décembre 2001]

Mireya Garcia-Vasquez [Bourse CONACYT]

Thomas Guionnet [Bourse INRIA]

Xavier Henocq [Bourse INRIA, jusqu'au 30 septembre 2001]

Reda Hosny [Bourse MAE]

Guillaume Larignon [Bourse INRIA, à partir du 1er octobre 2001]

Gaetan Le Guelvouit [Bourse INRIA-RÉGION]

Jean-Marie Pinel [Bourse MENRT]

Jérôme Viéron [Bourse INRIA]

2 Présentation et objectifs généraux

Les objectifs du projet sont de développer les concepts et les outils d'analyse, de modélisation, de codage, et de tatouage d'images, et plus généralement des informations vidéo manipulées en communication multimédia. Nos travaux portent plus particulièrement sur les problèmes suivants :

- l'interaction avec le contenu et la navigation dans des scènes vidéo 3D ;
- la représentation compacte et robuste aux bruits de transmission des images et des signaux vidéo ;
- le marquage (ou tatouage) des images et des signaux vidéo à des fins de protection contre les copies illicites, et à des fins d'authentification.

Pour répondre à ces objectifs dans un contexte en évolution forte et permanente, nos activités de recherche sont structurées en trois axes principaux :

- *Analyse et modélisation de séquences vidéo.* Dès que des fonctionnalités d'interaction avec le contenu vidéo ou de codage basé objet sont envisagées, une phase préalable d'analyse de la séquence vidéo est nécessaire. Cette phase consiste à segmenter les objets vidéo et à en extraire les principaux attributs (mouvement, texture, forme, orientation, illumination, ...). Les fonctions de navigation dans une scène vidéo nécessitent également la construction d'un modèle 3D de la scène. Les représentations des modèles et des attributs extraits doivent être faciles à manipuler. Lorsque des applications de communication sont envisagées, les représentations de ces modèles doivent en outre être compactes et robustes au bruit de transmission, rejoignant en cela le thème du codage conjoint source-canal. Pour une adaptation aisée à des contraintes variables liées au canal, ou pour une interaction fluide avec le contenu, il est en outre souhaitable que les modèles spécifiés soient bien adaptés à une description hiérarchique.
- *Codage conjoint source-canal.* Les environnements de communication hétérogènes (et à qualité de service non garantie) posent de nouveaux problèmes de représentation et de compression du signal image et vidéo. Les représentations codées d'une scène ou de ses attributs doivent être compactes et robustes au bruit de transmission. Le bruit de transmission peut se traduire par des pertes de paquets ou par des corruptions du train binaire. Minimiser l'impact du bruit sur la qualité du signal reconstruit, et optimiser la qualité globale de la chaîne de communication nécessitent d'abandonner les principes de séparabilité de codage de source et de codage de canal, pour se tourner vers des approches de codage conjoint source-canal. Nous nous intéressons aux techniques de codage robuste, de décodage robuste et conjoint source-canal mettant en œuvre le principe turbo (inspiré des turbo-codes). Nos travaux concernent également le problème de l'allocation de débit optimale entre le codage de source et le codage de canal, les techniques de codage scalable pour une adaptation aisée du débit de la source vidéo à la bande passante du réseau, et les techniques de codage par bases redondantes et par descriptions multiples.
- *Tatouage.* L'apparition de réseaux ouverts comme l'Internet suscite de nouveaux dangers pour la propriété intellectuelle. Grâce - ou à cause - des technologies numériques, il est

possible de faire des copies parfaites des contenus, la redistribution est ainsi devenue aisée et le traçage des redistributions difficile. Le problème du tatouage peut se formaliser comme un problème de communication d'information dans un environnement bruité. Les bases scientifiques de TEMICS, à la fois dans le domaine de l'analyse et du codage conjoint source-canal permettent ainsi d'appréhender de manière efficace cette problématique. Une autre spécificité de notre approche méthodologique est la prise en compte de l'impact mutuel du tatouage, de la représentation et de l'indexation des données au sein d'une base, notamment pour le traçage de copies illicites.

Compte-tenu de l'impact fort de la normalisation dans le domaine des applications multimédia communicantes, nous nous efforçons, dans nos orientations, de trouver un équilibre entre des approches compatibles avec les syntaxes normalisées et des approches se situant en rupture avec celles-ci. Le projet cherche également, en partenariat avec des industriels, à valoriser les résultats de ses travaux au travers de contributions à la normalisation (IETF, MPEG).

Les domaines d'application de nos travaux sont les applications multimédia en réseaux (Internet et réseaux sans fils) au travers de leurs multiples besoins en terme de compression, de transmission robuste ou en terme de fonctionnalités avancées telles que navigation, protection et authentification. Dans ces secteurs d'applications, nous avons des conventions de recherche ou de transfert avec France Télécom, le Celar, des projets multi-partenaires nationaux (projets RNRT COHRAINTE, COSOCATI, OSIAM, V2NET, VISI) et européens (projets IST-SONG, IST-OPENISE et IST-OZONE).

3 Fondements scientifiques

3.1 Modélisation 3D par géométrie projective

Mots clés : reconstruction 3D, vision par ordinateur, géométrie projective, projection perspective, modèle de caméra, équations de projection, auto-calibration, coordonnées homogènes, espace projectif, matrice fondamentale, matrice essentielle, contrainte épipolaire.

Résumé : *La modélisation (ou reconstruction) 3D désigne l'estimation de la forme et de la position d'objets 3D à partir d'un ensemble d'images de ces objets. TEMICS s'intéresse plus particulièrement au problème de la modélisation de grandes scènes à partir de séquences vidéo monoculaires. De par sa nature, la modélisation 3D par géométrie projective est un problème inverse. Nous nous concentrons sur l'un des problèmes non résolus de la reconstruction 3D par géométrie projective, à savoir sur l'estimation des paramètres 3D de la caméra, rendue particulièrement délicate lorsque cette estimation doit se faire à partir de nombreux points de vue distants fournis par une caméra mobile. Nous nous intéressons également à la cohérence des observations entre les points de vue, à la gestion des parties cachées et des discontinuités de profondeur.*

Les outils méthodologiques utilisés sont issus du domaine de la vision par ordinateur et de la géométrie projective. Les *équations de projection* pour une caméra C_i modélisée par une *projection perspective* s'écrivent :

$$\tilde{p}_i = P_i \tilde{x}, (1)$$

où \tilde{x} est un point 3D de coordonnées homogènes $\tilde{x} = (x \ y \ z \ 1)^t$ dans le repère de la scène \mathcal{R}_0 , et où $\tilde{p}_i = (X_i \ Y_i \ 1)^t$ représente les coordonnées de son projeté dans le plan image I_i . La matrice P_i est la *matrice de projection* associée à la caméra \mathcal{C}_i . Elle s'exprime en fonction de la matrice K représentant les *paramètres intrinsèques* de la caméra, et en fonction des transformations (rotation R_i et translation t_i) permettant de passer du repère de la scène \mathcal{R}_0 au repère de la caméra \mathcal{R}_i , par l'équation :

$$P_i = K(R_i|t_i). \quad (2)$$

La matrice K des *paramètres intrinsèques* de la caméra est définie par :

$$K = \begin{pmatrix} \frac{f}{l_x} & \frac{-f}{l_x \tan\theta} & X_c \\ 0 & \frac{f}{l_y \sin\theta} & Y_c \\ 0 & 0 & 1 \end{pmatrix},$$

où f est la *distance focale* de la caméra, et où l_x et l_y dénotent la taille en x et y des pixels. Le terme θ représente l'angle entre les axes x et y du plan image, et X_c et Y_c sont les coordonnées de la projection perpendiculaire du centre optique dans le plan image. Les matrices de rotation R_i et de translation t_i définissent les *paramètres extrinsèques* associés à la caméra \mathcal{C}_i .

Les paramètres intrinsèques et extrinsèques, aussi appelés paramètres de prise de vue, sont obtenus par des opérations de calibration ou d'auto-calibration. La *calibration* consiste à estimer les paramètres de prise de vue à l'aide de mires (objets fournissant des points 3D connus), et d'images de ces mires. L'*auto-calibration* consiste à estimer les paramètres de prise de vue uniquement à partir de données des images, préalablement mises en correspondance par identification et regroupement des projetés d'un même point 3D. Plus précisément, si on suppose les paramètres intrinsèques constants et connus, les paramètres extrinsèques R_i et t_i peuvent être estimés comme suit. Le repère de la scène est choisi arbitrairement confondu avec le repère \mathcal{R}_1 . Les matrices de rotation et de translation de la caméra \mathcal{C}_i sont donc celles qui font passer du repère \mathcal{R}_1 au repère \mathcal{R}_i de l'image I_i . Pour estimer ces matrices, on définit la *matrice essentielle* $E = T_i.R_i$ et la *matrice fondamentale* $F = K^{-t}.E.K^{-1}$. La matrice F est elle-même estimée à partir d'un ensemble de points mis en correspondance entre les images I_1 et I_i , par résolution du système formé de la *contrainte épipolaire* appliquée à chaque paire de points

$$p_1^t.F.p_i = 0. \quad (3)$$

La décomposition de E en valeurs singulières fait apparaître un produit de matrices qui permet de déduire la matrice de rotation R_i et le vecteur de translation t_i . Résoudre le problème de reconstruction 3D revient alors à trouver \tilde{x} connaissant les \tilde{p}_i , c'est-à-dire résoudre le système d'équations (1) relativement aux coordonnées \tilde{x} . Comme toute résolution de problème inverse, la reconstruction 3D est très sensible aux erreurs de mesure. Sa résolution requiert une bonne précision des mesures faites dans l'image, ainsi qu'un choix judicieux des techniques d'optimisation numérique.

3.2 Estimation de mouvement et segmentation

Mots clés : modèles affines, modèle de caméra, prédiction.

Résumé : *L'estimation du mouvement de la caméra, de régions ou d'objets au sein de séquences vidéo naturelles, ainsi que le calcul des segmentations correspondantes constituent des aspects essentiels sur lesquels s'appuient plusieurs axes de recherche du projet. La connaissance du mouvement permet en effet d'analyser le contenu dynamique d'une scène, ce qui facilite les manipulations du contenu objet des séquences ou sa reconstruction 3D. L'estimation de mouvement est un maillon très important dans les schémas de compression vidéo à base de prédiction temporelle, et dans les algorithmes de segmentation et de modélisation 3D.*

Le mouvement apparent directement visible dans les images représente le mouvement relatif entre l'objet et la caméra. L'estimation du mouvement propre des objets nécessite une compensation du mouvement de la caméra qui peut être réalisée en estimant le mouvement apparent d'une zone fixe dans l'image. L'estimation du mouvement peut être réalisée par pixel ou par région. Dans ce dernier cas, il est nécessaire de procéder à une segmentation des images considérées. TEMICS privilégie les approches par régions. Elles nécessitent l'utilisation d'une modélisation du mouvement 2D qui s'appuie souvent sur le modèle affine suivant :

$$\begin{aligned} dx &= t_x + k.(x - x_g) - \theta(y - y_g) + h_1(x - x_g) + h_2(y - y_g), \\ dy &= t_y + k.(y - y_g) + \theta(x - x_g) + h_1(y - y_g) + h_2(x - x_g), \end{aligned} \quad (4)$$

où t_x et t_y représentent les termes translationnels, et k le facteur de divergence. Le terme θ est l'angle de rotation plan, et h_1 et h_2 sont les termes hyperboliques. Il est important de noter que la vidéo ne restituant qu'une projection bidimensionnelle du mouvement tridimensionnel des objets, ce modèle ne permet d'obtenir qu'une représentation approximative du mouvement apparent 2D. Ce modèle est néanmoins souvent utilisé pour la compression, pour le recalage des images dans un référentiel de représentation commun dans le cadre de la construction d'images mosaïques 2D, ou encore pour la reconstruction 3D de la scène. Suivant le cas il peut être exploité sur des supports d'estimation de formes rectangulaires ou polygonales ou sur des supports issus d'une représentation par maillage triangulaire.

3.3 Approximation par maillages hiérarchiques et éléments finis

Mots clés : maillage, éléments finis, interpolation.

Résumé : *L'approximation par maillage et éléments finis constitue un outil méthodologique permettant une représentation souple et adaptative d'une image à des fins d'analyse ou de compression. Une image fixe peut être approximée par morceaux. Une séquence vidéo pourra être représentée par une succession de déformations appliquées à une représentation par maillage initiale.*

La méthode des éléments finis (MEF) peut être considérée comme une méthode générale de discrétisation et d'approximation de problèmes continus. Un milieu continu est subdivisé en un nombre fini d'éléments dont le comportement est régi par un nombre fini de paramètres. De façon pratique, la méthode des éléments finis est basée sur l'utilisation d'un maillage approprié du système à modéliser (les mailles usuelles étant généralement triangulaires ou quadrilatères).

Bien que destinée initialement à la modélisation et à l'analyse de problèmes physiques continus, une utilisation de la méthode des éléments finis dans le cadre du codage vidéo se justifie pour plusieurs raisons :

- la MEF offre une représentation souple et adaptative, elle se prête ainsi parfaitement au caractère non stationnaire du signal vidéo en espace et en temps, en alternant zones de luminance, de mouvement, uniformes et/ou perturbées,
- la MEF permet une analyse numérique robuste et, par analogie avec la physique, permet une large compréhension des problèmes et des solutions envisageables.

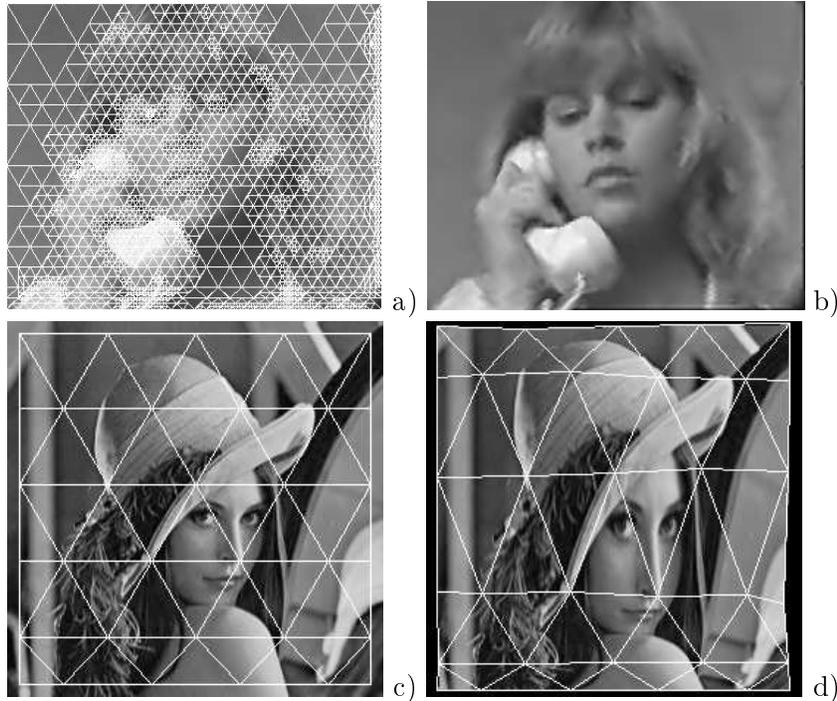


FIG. 1 – Exemples d'utilisation des maillages. (a,b) : exemple d'utilisation d'un maillage pour représenter une image fixe. (c,d) : exemple d'utilisation d'un maillage pour déformer une image.

Le type de modèle utilisé sur les mailles peut être de nature différente suivant la complexité ou bien encore le type de continuité désiré (éléments finis de Lagrange réalisant une interpolation affine par maille et de continuité C^0 , éléments finis de Hermite réalisant une interpolation par des polynômes de degré 3 par maille et de continuité C^1). L'interpolation de Lagrange pour un maillage triangulaire peut être définie par la donnée d'un ensemble de valeurs définies sur les nœuds du maillage, et l'interpolation de ces valeurs de façon affine sur les différentes mailles. Ainsi, si l'on considère un triangle $(P_1P_2P_3)$, avec des valeurs aux nœuds (V_1, V_2, V_3) , l'interpolation affine pour tout point de ce triangle s'exprime comme :

$$V(P) = w_1(P).P_1 + w_2(P).P_2 + w_3(P).P_3$$

avec

$$w_1(P) = \frac{\overrightarrow{P_2P} \wedge \overrightarrow{P_2P_3} \cdot \vec{z}}{\overrightarrow{P_2P_1} \wedge \overrightarrow{P_2P_3} \cdot \vec{z}},$$

où \wedge et \cdot représentent respectivement le produit vectoriel et le produit scalaire de deux vecteurs, et \vec{z} , le vecteur orthogonal au plan de l'image. Les poids $w_2(P)$ et $w_3(P)$ sont obtenus de façon similaire par permutation circulaire sur les indices.

Dans le cadre d'une image fixe, l'image est approximée par morceaux de façon continue sur les différentes mailles (cf. figure 1.a,b). Dans le cas d'une image issue d'une séquence, le maillage est utilisé afin de déformer une image de référence dans le but d'obtenir une image aussi proche que possible de l'image courante (cf. figure 1.c,d). Les différents paramètres contrôlant le maillage sont alors obtenus selon des critères d'optimisation de la reconstruction de l'image désirée.

3.4 Théorie débit-distorsion et formalisme MDL

Mots clés : Limite OPTA (Optimum Performance Theoretically Attainable), Formalisme MDL (Minimum Description Length), allocation de débit, optimisation débit-distorsion, codage avec pertes, codage par descriptions multiples, modélisation de canal.

Résumé : *Les travaux de codage (avec ou sans pertes) et les travaux de codage conjoint source-canal de TEMICS reposent sur des concepts fondamentaux de la théorie de l'information, comme les notions de quantité d'information, d'entropie d'une source avec et sans mémoire, de capacité d'un canal et de limite de performance débit-distorsion. Les algorithmes de compression sont conçus de manière à approcher au mieux la limite de performance débit-distorsion $R(D)$ pour un signal considéré. Le coût de description d'un message peut aussi être représenté à l'aide du formalisme MDL (minimum description length) qui se rapproche des notions de complexité de Kolmogorov et de complexité stochastique.*

Le théorème du codage de source définit les limites de performances du codage sans pertes et avec pertes. Dans le cas du codage sans pertes, la borne inférieure de débit est fournie par l'entropie de la source. Dans le cas du codage avec pertes, elle est donnée par une fonction débit-distorsion $R(D)$. Cette fonction $R(D)$ établit la quantité d'information minimale nécessaire à la représentation du message pour une distorsion donnée de ce message, ou à l'inverse, établit la distorsion minimale pour un débit (ou une quantité d'information) donné.

Cette fonction débit-distorsion limite, appelée OPTA (*Optimum Performance Theoretically Attainable*), est en général difficile à déterminer de manière analytique, sauf pour le cas de sources particulières (théoriques), comme par exemple pour des sources Gaussiennes. Dans le cas de sources réelles (images ou signaux vidéo), les algorithmes de compression nécessitent l'utilisation de procédures d'optimisation débit-distorsion. Pour ce faire, une minimisation Lagrangienne est généralement mise en œuvre :

$$\frac{\partial J}{\partial Q} = 0 \text{ o } J = R + \lambda \text{ avec } \lambda > 0.$$

Cette optimisation est réalisée en dérivant le Lagrangien, J , par rapport au paramètre à l'origine de la distorsion, c'est-à-dire souvent par rapport au pas de quantification Q .

Dans un cadre d'optimisation des performances de la qualité de service d'un système de communication, il est en outre nécessaire de prendre en compte les caractéristiques du canal et du codage de canal, en développant des stratégies de codage conjoint source-canal. On cherche alors à minimiser une distorsion globale, définie comme la somme de la distorsion due à la compression du signal et de la distorsion induite par le bruit de transmission sur le canal, sous contrainte d'un débit global donné. La fonction de coût à minimiser devient alors

$$J = (R_s + R_c) + \lambda(D_s + D_c).$$

Le débit global est défini comme la somme du débit de source R_s et du débit de redondance R_c introduit par le codeur de canal (ou codes correcteurs). Les techniques d'optimisation débit-distorsion conjointes source-canal permettent d'améliorer le compromis compression - robustesse au bruit de transmission, d'avoir un partage optimal de la bande passante d'un canal de communication entre le débit utile alloué à la source et le débit de redondance alloué à la protection. La redondance nécessaire à la protection du signal peut être générée à l'aide de codes correcteurs traditionnels ou alors être introduite dans la représentation compressée du signal au travers de transformations par bases redondantes (voir section 3.5), ou encore par des approches de représentation du signal par descriptions multiples. Le *codage par descriptions multiples* peut être vu comme une généralisation à un environnement de transmission multi-canal du codage avec critère de fidélité.

Le formalisme MDL (Minimum Description Length) constitue un outil méthodologique supplémentaire pour quantifier le coût de description global, ou la quantité d'information minimale, nécessaire pour représenter un message. Cet outil se rapproche de la notion de quantité d'information et de complexité stochastique définie par Kolmogorov. Il est basé sur une modélisation du signal par une suite de réalisations de variables aléatoires. En notant s une chaîne d'observations des réalisations de ces variables, et θ le vecteur des paramètres du modèle de représentation, le principe du formalisme MDL est de minimiser le coût de description de la chaîne s :

$$MDL(s) = \min_{\theta} \{-\log_2 P_{\theta}(s) + DL(\theta)\},$$

où $DL(\cdot)$ représente le coût de description. Le terme $-\log_2 P_{\theta}(s)$ correspond au coût de codage idéal de la chaîne s connaissant le modèle de représentation θ ; $-\log_2 P_{\theta}(s) + DL(\theta)$ est le coût de description global de la chaîne s en utilisant le modèle θ . Le formalisme MDL est souvent utilisé dans des schémas de codage d'images fixes et animées, afin d'optimiser le modèle de représentation de l'image.

3.5 Représentation temps-fréquence et par bases de fonctions redondantes

Mots clés : Transformations par ondelettes, bancs de filtres multi-fréquences, échantillonnage, codes correcteurs, descriptions multiples.

Résumé : *Les algorithmes de compression du signal présentent toujours plus ou moins la même architecture : une première étape de décorrélation suivie d'opérations*

de quantification et de codage. Les transformations (transformations en ondelettes ou par paquets d'ondelettes) qui remplacent le signal d'origine par sa représentation sur une base de fonctions, correspondent à une telle approche et constituent des éléments essentiels des schémas de codage développés par TEMICS. Ces transformations constituent en effet des outils privilégiés pour une représentation scalable des signaux compressés, et donc pour permettre des fonctions de régulation de débit dans des environnements de communication hétérogènes. Dans un contexte de transmission sur des canaux bruités, il s'avère nécessaire de maintenir une certaine corrélation ou redondance dans la représentation compressée du signal. L'une des approches étudiées par TEMICS repose sur l'utilisation de transformations sur des bases de fonctions redondantes.

Les techniques de représentation temps-fréquence (ou multi-résolution) du signal consistent à décomposer le signal par bancs de filtres multi-fréquences. Une structure itérée de décomposition par bancs de filtres va permettre de créer un ensemble de sous-signaux (aussi appelés sous-bandes) qui vont contenir l'information relative aux différentes bandes de fréquences contenues dans l'image ou le signal vidéo, comme illustré sur la figure 2. Ces structures de décomposition, conduisant naturellement à une représentation multi-échelle et scalable du signal, occupent une place de plus en plus importante dans les systèmes de compression et de communication. Elles constituent en effet des outils privilégiés pour la mise en place de mécanismes de contrôle de congestion et de régulation de débit dans des environnements de communication hétérogènes, et en particulier multi-points.

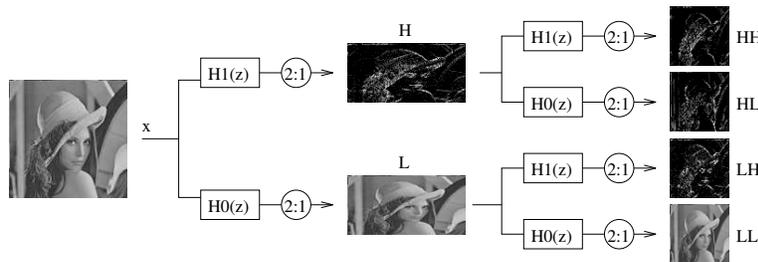


FIG. 2 – Analyse multi-résolution d'un signal 2D par un banc de filtres : le signal est décomposé sur 4 espaces orthogonaux entre eux, un espace d'approximation (correspondant au signal LL contenant les basses fréquences verticales et horizontales) et trois espaces de détail (correspondant aux signaux LH , HL et HH).

Les décompositions par bancs de filtres multi-fréquences utilisées en compression de signal sont à échantillonnage critique, c'est-à-dire que la somme des échantillons contenus dans les sous-signaux est égale au nombre de pixels contenus dans l'image. Dans un contexte de transmission sur des canaux bruités, il s'avère nécessaire de maintenir une certaine redondance ou corrélation dans la représentation multi-résolution de l'image. L'une des approches considérées dans le projet TEMICS consiste à utiliser des transformations redondantes, comme par exemple des bancs de filtres sur-échantillonnés. Conceptuellement, l'approche s'apparente à un code de canal en bloc, défini sur le corps des réels, et mis en œuvre au travers de la transfor-

mation. Les paramètres n/m définissant le taux de rendement du code et les fonctions de base de la transformation peuvent être adaptés aux caractéristiques du canal à un instant donné. Ces transformations peuvent être vues comme une décomposition du signal sur un ensemble de fonctions de bases $\Phi = \{\phi_k\}_{k=1}^m$ redondantes et être formalisées par une transformation linéaire $F : C^n \rightarrow C^m$ d'un espace de dimension n vers un espace de dimension $m \geq n$. L'interprétation de ces transformations redondantes sous forme de codes de canal permet d'appliquer certains outils de décodage de canal (comme par exemple les algorithmes de décodage basés syndrome ou des estimateurs Bayésiens) pour exploiter la redondance et corriger les erreurs (ou pertes) dues à la transmission.

3.6 Tatouage d'images

Mots clés : tatouage (watermarking), signature, copyright, authentification, détection, extraction robuste, étalement de spectre.

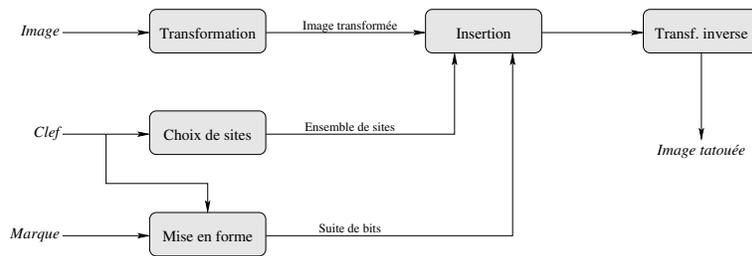
Résumé : *L'objectif général du tatouage (watermarking en anglais) est de dissimuler des informations dans une image afin de pouvoir identifier son origine, de protéger son copyright, ou bien encore de vérifier son intégrité. Cette marque doit être indécélable, infalsifiable et ineffaçable. Son extraction doit être robuste face à divers types d'attaques que l'image pourrait subir (compression, retaille, filtrage, transformations géométriques, ...). Il est possible d'établir un parallèle entre le tatouage et une situation de communication classique. TEMICS privilégie le marquage dans le domaine transformé, qui s'apparente en communication aux techniques de transmission multi-porteuses. Cette approche permet en particulier d'exploiter des fonctions de masquage du système visuel afin de s'assurer que la marque sera non visible.*

Les techniques d'insertion de marques peuvent être classées en plusieurs catégories : marquage de sites choisis dans une image (dans le domaine spatial), marquage dans le domaine transformé (par exemple après transformation en cosinus discrète), ou par utilisation de codes pseudo-aléatoires pour répartir la marque dans tout le signal. Le schéma général d'insertion d'une marque dans le domaine transformé et d'extraction de cette marque est illustré par la figure 3.

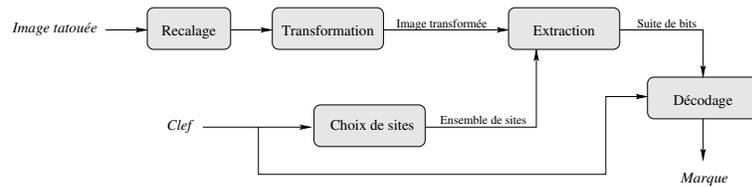
Une fois le site d'insertion choisi, il faut concevoir une technique d'insertion de la marque. Il existe plusieurs méthodes d'insertion. Nous nous intéressons principalement aux techniques d'insertion par étalement de spectre, qui offrent une meilleure robustesse aux attaques que peut subir la marque.

Soit un message m à insérer défini par la suite de bits $\{b_j, j = 1, \dots, M\}$, et soit $\{V_i, i = 1, \dots, N\}$ l'ensemble des valeurs des coefficients correspondant aux différents sites d'insertions. La technique d'étalement de spectre consiste alors, à partir d'un jeu de vecteurs $\{g^j, j = 1, \dots, M\}$ de dimension N aléatoires et orthogonaux (par exemple défini à partir de M-séquences), à insérer le message en modifiant les valeurs originales des coefficients :

$$V'_i = V_i + \alpha * \sum_{j=1..M} b_j \cdot g_i^j.$$



(a) Schéma du principe d'insertion de la marque.



(b) Schéma du principe d'extraction de la marque.

FIG. 3 – Schéma de principe pour l'insertion et l'extraction de marque dans une image.

Par la suite, on peut extraire le bit b_k par corrélation : $\frac{\sum_{i=1..N} V_i' \cdot g_i^k}{\sum_{i=1..N} (g_i^k)^2}$, puisque les vecteurs g^j sont orthogonaux entre eux, et qu'ils sont décorrélés de l'image. Le coefficient α permet de régler l'amplitude du bruit généré dans l'image par l'insertion de la marque. Ce coefficient doit être choisi suffisamment faible afin que la marque ne soit pas visible, mais aussi suffisamment fort pour que la détection par corrélation fonctionne correctement même en présence d'attaques.

4 Domaines d'applications

4.1 Panorama

Nous nous intéressons à trois grands types d'applications. Le premier concerne la compression d'images fixes et animées. Nous explorons dans ce cadre les domaines de la compression scalable du signal vidéo (en collaboration avec France Télécom R&D et dans le cadre du projet RNRT VISI), et le couplage de la fonction de compression à des fonctionnalités telles que la navigation, l'indexation et le tatouage. Le second secteur concerne les applications de télécommunications incluant les applications de consultation multimédia (ex : visite virtuelle de sites), de diffusion (diffusion audiovisuelle sur Internet) et de communications de groupes (ex : visioconférence, réunions virtuelles). Nous avons dans ce domaine d'application de nombreuses collaborations nationales (cinq projets RNRT) et Européennes (deux projets IST). Le troisième secteur a trait à la protection et à l'enrichissement des contenus visuels (images fixes et vidéo) par insertion de données numériques dans le signal (tatouage). Enfin, certains travaux

menés dans l'équipe peuvent avoir des retombées dans les domaines de la production, de la post-production vidéo et de la structuration de séquences vidéo pour de l'aide à la navigation, même si ces derniers domaines ne constituent pas des domaines d'application privilégiés par le projet.

4.2 Compression d'images fixes et animées avec fonctionnalités avancées

La compression d'image et du signal vidéo constitue un cadre applicatif naturel d'un grand nombre de techniques que nous développons. La compression occupe en effet une place prépondérante dans de nombreux domaines d'application nécessitant un stockage de l'information image ou vidéo, ou une transmission de cette information sur des réseaux à bande étroite, voire à qualité de service non garantie. Néanmoins, nos travaux ne cherchent pas à répondre seulement à des objectifs de compression mais cherchent à allier la compression à des fonctionnalités nouvelles telles que la navigation dans une scène par reconstruction de points de vues intermédiaires, la transmission fiable de ces données sur des réseaux à qualité de service non garantie et hétérogènes, ou encore telles que la protection et l'enrichissement des contenus par insertion de données numériques dans le signal (tatouage).

Ainsi, nos travaux algorithmiques d'analyse, de représentation par maillages (et en particulier par maillages hiérarchiques), de modélisation 3D de scènes réelles, et de création de vues intermédiaires permettent d'allier des fonctions de compression et de navigation dans des scènes virtuelles 3D issues de séquences vidéos naturelles. Ils s'appliquent donc directement à toute application multimedia mettant en œuvre des scènes virtuelles 3D, par exemple visite de musées virtuels, télé-présence, jeux,... Ce domaine d'application fait l'objet du projet RNRT-V2net.

De manière générale, le besoin d'interactions avec le contenu vidéo, d'accès et de navigation au sein du contenu demande de structurer l'information vidéo (objets, mosaïques, plans, ...) et de mettre au point des représentations compactes de cette structuration. Il semble donc naturel de chercher à concevoir des schémas de compression générant une information qui peut être aisément exploitée lors d'une structuration de la séquence vidéo et à l'inverse exploitant de manière efficace l'information de description ou de structuration (qui peut se présenter sous forme de métadonnées) à des fins de compression. L'accès à des bases de données vidéo structurées constitue ainsi un cadre applicatif naturel de nos travaux en compression avec fonctionnalités avancées.

4.3 Communication multimédia

Les applications diffusant du contenu multimédia sont confrontées aux problèmes de transmettre sur des réseaux aux performances variables des flux volumineux, et selon un modèle plus complexe que le traditionnel point-à-point homogène, comme par exemple en multi-point dans des environnements hétérogènes fixes ou mobiles. Pour réussir à transmettre les flux sur les réseaux, ces applications ont traditionnellement recours aux techniques relevant du domaine de la compression. Or les outils définis à des seules fins de compression présentent certaines limitations : ils sont très sensibles aux pertes ou aux erreurs de transmission. Nos travaux en codage conjoint source-canal ont pour objectifs de fournir des solutions au problème de l'opti-

misation de la qualité de service dans des environnements offrant un service de transport non garanti (comme par exemple, l'Internet d'aujourd'hui). Ils s'appliquent ainsi et de façon naturelle aux applications multimédia communicantes dans de tels environnements (Internet, liens sans fils). La transmission d'images ou de signaux vidéo dans des environnements hétérogènes (fixes et mobiles) en multi-point, constitue également un domaine d'application des travaux de recherche que nous menons. En particulier, la mise au point de schémas de codages scalables nous permet d'offrir des techniques élaborées de régulation de flux pour faire du contrôle de pertes et de congestion dans des contextes de diffusion multi-points.

Les domaines applicatifs plus particulièrement traités par le projet, au travers de ses partenariats industriels bilatéraux et européens, concernent tout d'abord les applications de consultation multimédia sur Internet (télé-achat, télé-visite, télé-enseignement). Dans ce type d'application, le codage des contenus multimédia (audiovisuels, 3D synthétique) est en général réalisé en différé, ce qui conduit à des difficultés supplémentaires pour assurer un contrôle de congestion efficace. Cette caractéristique va ainsi fixer un cadre spécifique pour une partie de nos travaux. Ce secteur applicatif constitue un point de rencontre avec les domaines de l'indexation, de la consultation et de la gestion de bases de données vidéo. Nos travaux concernent également les communications de groupes (ex : visioconférence) que nous traitons en collaboration avec France Télécom R&D et au sein du projet RNRT Visi. Enfin ils concernent les vidéocommunications mobiles, domaine d'application que nous abordons en collaboration avec le Celar au travers d'une application de téléguidage d'un robot terrestre mobile, et au sein du projet RNRT Cosocati.

4.4 Protection, authentification et enrichissement de contenus images et vidéo

Les technologies numériques, qui permettent l'apport de fonctionnalités nouvelles dans les applications de communication vidéo et multimédia, ainsi que l'apparition de réseaux ouverts comme l'Internet ne sont pas sans susciter de nouveaux dangers pour la propriété intellectuelle et les droits de diffusion. De l'avis des professionnels de la photographie, le problème principal lié à la mise à disposition de photos au travers de réseaux ouverts est le pillage illicite de leur patrimoine. Il est en effet possible de faire des copies parfaites des contenus, la redistribution est devenue aisée et le traçage des redistributions est difficile. Ce problème, exprimé dans un contexte informatique, demande de mettre en place un environnement permettant d'une part de détecter qu'une image ou une vidéo fassent illégalement partie d'une collection d'images mises à disposition par un tiers (traçage de copies) et d'autre part d'exhiber une preuve de propriété irréfutable (authentification). Ces problèmes de vérification de propriété intellectuelle, de traçage d'usage illicite, et d'authentification des images numériques constituent ainsi un domaine d'application à forts enjeux de nos études de marquage (aussi appelé tatouage ou " watermarking " en anglais).

Pour atteindre ces deux objectifs, il peut s'avérer utile de combiner les techniques de tatouage et d'indexation par le contenu. Cette combinaison peut en effet permettre de couvrir les deux aspects que sont traçage et authentification, mais aussi de renforcer la puissance et la robustesse de chaque technique prise individuellement car l'une et l'autre s'épaulent dans le processus de recherche et de vérification. Ces domaines d'application vont être couverts dans

le cadre du nouveau projet RNRT Diphonet qui va démarrer en janvier 2002.

Néanmoins, la protection de copyright, l'authentification et le traçage de copies illicites ne sont pas les seuls domaines d'application du tatouage. Les outils d'insertion de données numériques dans un signal vidéo peuvent être utilisés pour transporter des métadonnées dans un environnement de diffusion hétérogènes, assurant ainsi une certaine transparence du transport de ces métadonnées et une certaine persistance de ces données dans un cadre de transcodage inhérent à l'application de diffusion sur une infrastructure hétérogène de transport (DVB, Internet, accès mobiles, etc...) ou vers un parc hétérogène de terminaux. Cela conduit alors à des contenus enrichis par des métadonnées enfouies dans le signal, lesquelles données peuvent être exploitées de façon transparente par le terminal pour offrir différentes fonctionnalités à valeur ajoutée.

5 Logiciels

5.1 Plate-forme de communication vidéo

Participants : Bilal El Ali, Christine Guillemot [correspondant], Reda Hosny, Jérôme Viéron.

Une plate-forme logicielle de communication multimédia a été développée. Celle-ci, en émulant une gamme de qualité de service réseau, devrait permettre de démontrer, d'expérimenter et de valider nos algorithmes de représentation, de codage et de codage conjoint source-canal dans des conditions *réelles* de transmission. Elle intègre les protocoles de transmission multimédia (RTP/RTCP), des mécanismes de différenciation de services, et des mécanismes de contrôle de congestion de bout-en-bout, à la fois en point-à-point (Unicast) et multi-point (Multicast).

5.2 Codecs vidéo et interface réseau

Participants : Philippe Coval, Jeanne Guillou, Luce Morin, Stéphane Pateux [correspondant].

Pour valider nos travaux algorithmiques de compression et de codage conjoint, nous avons développé, optimisé et intégré à la plateforme de communication deux codecs (codeurs-décodeurs) vidéo. Un premier codec vidéo scalable, basé sur la norme H.263+, intègre des outils de codage robuste au bruit de transmission développé au sein du projet. Deux modes de représentation scalable sont mis en œuvre : la scalabilité spatiale et en rapport signal à bruit (dite scalabilité SNR). Un second décodeur vidéo repose sur l'algorithme de compression basé modèles 3D décrit en section 6.1.1. Ces deux décodeurs ont été interfacés à la plateforme de communication décrite ci-dessus via un lecteur (ou *player*) générique de flux qui permet de décoder et de restituer à *la volée* les flux vidéo diffusés sur le réseau.

6 Résultats nouveaux

6.1 Analyse et modélisation de séquences d'images

Mots clés : mouvement, champs denses, disparité, segmentation, suivi temporel, triangulation, maillages 2D et 3D, maillages actifs, maillages hiérarchiques, contour actif, cartes de profondeur, géométrie projective, modèles par facettes, réalité augmentée, ombrage, modèles d'illumination, image mosaïque.

Résumé : *Pour répondre à des besoins de navigation dans une scène, au travers d'un réseau de télécommunications, nous avons développé un schéma global de reconstruction et de représentation d'une séquence vidéo par un ensemble de modèles 3D. La représentation par modèles 3D permet la reconstruction de vues intermédiaires. Dans cette étude nous considérons des grandes scènes complexes et de longues séquences d'images, dont l'acquisition est faite par une caméra unique et mobile. Une autre spécificité de l'étude repose sur le fait que nous ne disposons pas d'information a priori sur le contenu de la scène, ni sur les paramètres de calibration de la caméra. Cette étude démarrée en 2000 s'est poursuivie cette année en cherchant à résoudre quelques points durs tels que la prise en compte des occultations dans l'estimation du mouvement ou encore la gestion des discontinuités entre les modèles. Les travaux de l'année ont aussi porté sur le développement d'un algorithme de segmentation et de codage conjoint par maillage actifs. Les maillages actifs sont aussi exploités pour la création de mosaïques généralisées d'objets déformables.*

6.1.1 Modélisation 3D par analyse de séquences vidéo monoculaires

Participants : Franck Galpin, Luce Morin.

Les approches de modélisation, basées sur la mise en correspondance de points de vue distants dans les séquences, reposent sur des techniques d'estimation conjointe du relief de la scène observée et du mouvement de la caméra. En l'absence de données a priori, le mouvement de la caméra est déterminé à partir de la matrice fondamentale, elle-même estimée à partir de points singuliers extraits et suivis dans la séquence. L'estimation du mouvement dans la séquence est une étape cruciale de la modélisation 3D.

Nos travaux cette année ont d'abord porté sur la prise en compte des occultations dans l'estimation du mouvement par maillage triangulaire. Nous avons montré que, lorsque les paramètres internes de la caméra ne sont connus qu'approximativement, les méthodes classiques de décomposition de la matrice fondamentale aboutissent à des solutions incohérentes avec la géométrie épipolaire estimée. Nous avons alors proposé une méthode robuste d'estimation du mouvement de la caméra, basée sur un algorithme itératif intégrant un calcul de pose par la méthode de Dementhon [25]. Nous avons ensuite amélioré notre schéma global de reconstruction-représentation d'une séquence par création d'un ensemble de modèles 3D. L'originalité de l'approche est de ne pas reconstruire un modèle 3D unique de la scène, mais une suite de modèles 3D indépendants liés par des points de vue communs (images clés). Ce

schéma génère un flux séquentiel de modèles VRML pouvant facilement être transmis sur un réseau en vue d'une navigation à distance dans la scène. Cependant du fait de l'indépendance des modèles, certaines fonctionnalités apportées par une modélisation 3D telles que l'incrustation d'un objet dans une séquence, ne sont plus réalisables car elles requièrent un référentiel constant. Nous avons alors développé un algorithme d'ajustement par fenêtre glissante (*sliding bundle adjustment*) produisant à la volée un recalage géométrique des modèles 3D successifs, sous la contrainte de continuité de la séquence reconstruite.

La méthode de modélisation a été testée et validée sur de nombreuses séquences réelles. Elle a été appliquée à des manipulations de scène telles que le ré-éclairage, l'insertion d'objets avec occultation (voir figure 4), la stabilisation de séquences et la génération de séquences stéréoscopiques sur matériel spécialisé (salle de Réalité Immersive de l'IRISA). Des démonstrations sont disponibles sur le site web de l'équipe (<http://www.irisa.fr/temics/Demos/3D4>).

6.1.2 Segmentation d'objet vidéo pour le codage

Participants : Marc Chaumont, Dubhe Chavira-Martinez, Henri Nicolas, Stéphane Pateux.

Les travaux en codage orienté-objets reposent souvent sur une segmentation de la scène qui se veut être aussi fidèle que possible aux objets physiques présents. Dans le cadre du projet RNRT Osiam, nous avons développé une technique de segmentation et de suivi temporel répondant à un tel critère. Les développements de cette année ont porté sur l'introduction d'une métrique statistique spatio-temporelle afin d'améliorer le schéma de suivi, notamment la gestion des zones indéterminées liées aux occultations entre objets.

Toutefois, dans le cadre d'une application de codage, la segmentation idéale ne correspond pas forcément à une segmentation sémantique (un exemple typique concerne les objets immobiles qui n'ont pas besoin d'être segmentés). Nous avons alors développé un nouvel algorithme de segmentation mieux adapté à des objectifs de codage et de compression. Les objets vidéos recherchés sont définis comme ayant une texture variant peu au cours du temps, et ayant un mouvement continu. Pour chaque objet, un modèle de déformation par maillages actifs est utilisé. Un suivi utilisant des mesures de cohérence à long-terme de ces maillages a été mis en place. Le problème de la segmentation est alors posé comme un problème d'étiquetage minimisant une fonctionnelle énergétique (modèle Markovien). Les termes énergétiques et les modèles utilisés sont adaptés à des objectifs de compression.

6.1.3 Représentation par maillages et mosaïques généralisées d'objets vidéo

Participants : Marc Chaumont, Stéphane Pateux.

L'outil de représentation par maillage déformable hiérarchique des objets dans une séquence est utilisé comme support d'estimation et de compensation du mouvement d'un objet dans une séquence vidéo. Afin de permettre un suivi à long terme du maillage, cet estimateur a été robustifié en introduisant une technique d'estimation de mouvement multi-grille. L'outil de maillage déformable hiérarchique couplé à l'estimateur de mouvement est utilisé pour plaquer par suivi à long terme et compensation la texture de l'objet observé sur une image de référence.



a)



b)



c)



Cet outil est utilisé à la fois dans un contexte de codage et de création d'images mosaïques (voir fig. 5). Les techniques de création d'images mosaïques sont souvent basées sur des hypothèses fortes ; les mouvements de la caméra doivent appartenir à une certaine classe (en général mouvement de type '*pan & tilt*'), à défaut, le contenu de la scène doit être à une profondeur éloignée (afin de pouvoir utiliser un modèle de mouvement global - souvent homographique ou quadratique). Par ailleurs la scène observée doit être une scène rigide. L'utilisation de maillages actifs permet d'éviter au moins en partie ces hypothèses et de créer des images mosaïques pour des objets déformables. L'image mosaïque générée peut être à contenu temporellement dynamique afin de refléter les variations temporelles de texture. L'utilisation d'un filtre de Kalman réglable peut permettre de diminuer le bruit de la caméra.

6.1.4 Analyse des conditions d'illumination dans une séquence d'images

Participants : Henri Nicolas, Jean-Marie Pinel.

Les applications de compression, d'édition de séquences et de réalité augmentée peuvent bénéficier d'outils de segmentation des zones d'ombres dans la séquence. Cette segmentation nécessite d'une part de caractériser en intensité et en position les sources de lumière (supposées ponctuelles) présentes dans la scène. La position de la source de lumière est estimée par analyse des contours de l'objet (après segmentation) et de la position d'un cône d'ombre associé à cet objet. La segmentation de l'ombre est ensuite affinée en minimisant un critère combinant la position estimée de la source de lumière, la distance au contour de l'objet, les variations d'illumination et les caractéristiques texturales de la zone d'ombre à estimer. Le résultat est illustré sur la figure 6.

La création d'images mosaïques nécessite aussi une certaine harmonisation de l'éclairage des images *fusionnées*. Le coefficient de recouvrement (ou coefficient de "blending") est défini de manière adaptative dans la zone de recouvrement de deux images. La zone de recouvrement est découpée en deux régions délimitées par une "ligne de partage". Cette ligne de partage est calculée de manière à ce que les variations de texture de part et d'autre de cette ligne soient la plus faible possible. Une compensation d'illumination globale est alors appliquée sur une des images en fonction des variations d'illumination résiduelles de part et d'autre de la ligne de partage. Cette technique permet d'améliorer la qualité des images mosaïques lorsque les images originales ont des illuminations significativement différentes.

6.1.5 Structuration par analyse des trajectoires des objets vidéo

Participants : Franck Denoual, Mireya Garcia-Vasquez, Henri Nicolas.

Les segments d'une séquence vidéo peuvent être caractérisés par le mouvement des objets présents entre deux instants clés. Les instants clés sont déterminés par des changements d'orientation d'un objet par rapport à la caméra. Ces changements peuvent être dus à des mouvements de rotation non plan ou de translation d'amplitude forte. Les objets supposés rigides sont d'abord segmentés. Le mouvement apparent (avec ses paramètres de translation, de rotation 2D et son facteur d'échelle) est estimé entre deux instants clés. La validité du modèle de mouvement plan est évaluée à l'aide d'un critère prenant en compte la qualité de compensation

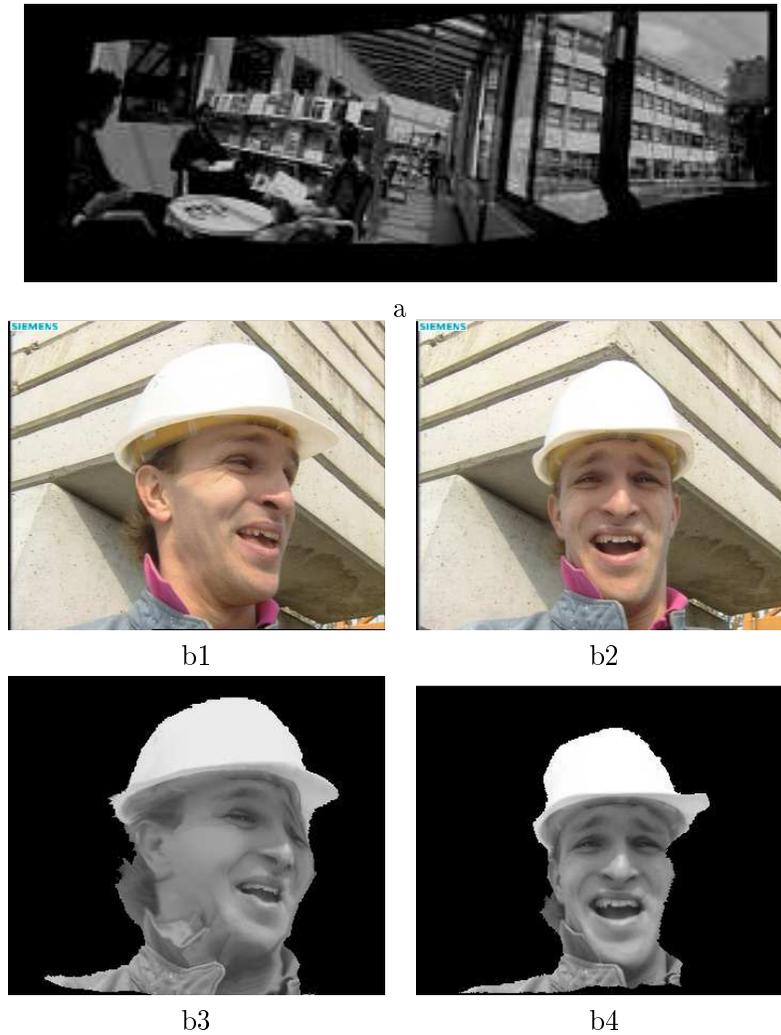


FIG. 5 – Exemple de mosaïque d'objet vidéo. a) scène d'intérieur avec un mouvement de la caméra en présence d'objets proches. b) image mosaïque d'un objet déformable (b1-b2 : images originales, b3 : mosaïques pour la personne en avant plan, b4 : image reconstruite à partir de la mosaïque).



FIG. 6 – Segmentation des ombres.

de mouvement, ainsi que les dimensions et la structure géométrique des zones non-recouvertes. Lorsque ce modèle n'est pas valide, cela signifie que l'objet a changé d'orientation par rapport à la caméra. La direction de l'axe de rotation de l'objet et l'angle de rotation correspondant sont alors déterminés en maximisant un critère basé d'une part sur la détection des zones découvertes et recouvertes de l'objet, et des zones pour lesquelles la mise en correspondance obtenue avec le modèle de mouvement plan estimé n'est pas satisfaisante. Finalement, l'intervalle de temps considéré est divisé en deux segments et le processus est réitéré jusqu'à ce que l'on détermine les instants correspondants au début et à la fin du mouvement générant les variations d'orientation.

6.2 Codage et codage conjoint source-canal

Mots clés : compression, codage et décodage robustes, contrôle de congestion, descriptions multiples, scalabilité, transformations en ondelettes, multi-ondelettes, transformations par bases de fonctions redondantes, quantification, théorie débit-distorsion, régulation, progressivité, codes correcteurs, modélisation stochastique, estimation Bayésienne, inférence probabiliste, propagation de croyance, principe turbo, modèles de canal, Internet, mobilité, qualité de service.

Résumé : *Les travaux ont d'abord porté sur le codage à bas débit et scalable à grain fin de séquences vidéo. La première approche repose sur une représentation de la séquence par une suite de modèles 3D de la scène. La seconde approche*

est basée sur une représentation multi-résolution spatio-temporelle suivi d'un codage EBCOT. Deux variantes portant sur la structure (maillage triangulaire, ou "quadtree" contraint en débit) sous-jacente à l'estimation et à la compensation du mouvement ont été mises en œuvre. Nous avons continué nos travaux en contrôle de congestion et de débit des flux vidéo en nous concentrant sur les transmissions multi-points. Plusieurs approches de codage robuste à la fois à des pertes de paquets et à des perturbations binaires ont par ailleurs été étudiées. Rappelons que le cadre applicatif de ce travail est la transmission du signal vidéo sur des environnements hétérogènes tels que l'Internet filaire ou sans fils. Les approches de codage robuste développées sont basées sur l'utilisation de bases de fonctions redondantes et de techniques de codage par descriptions multiples. Dans le domaine des communications sur des liens sans fils, en collaboration avec le projet SIGMA2, une modélisation sous forme d'automate stochastique d'un codeur statistique a permis de développer un algorithme de décodage robuste de codes à longueurs variables. Un mécanisme de synchronisation douce de codes à longueurs variables permettant de réduire les phénomènes de désynchronisation dont souffrent ces codes en présence de bruit de transmission a en outre été mis en œuvre. Ces principes de décodage sur treillis et de synchronisation sont actuellement étendus aux codes arithmétiques qui tendent à occuper une place de plus en plus prépondérante dans les schémas de compression émergents.

6.2.1 Codage orienté modèles 3D

Participants : Franck Galpin, Luce Morin.

Nous avons poursuivi l'étude du schéma de compression vidéo basé sur une modélisation 3D de la scène. Ce schéma est destiné à coder à très bas débit des scènes fixes. Le schéma développé suit une structure I-P-B (Intra-Prédit-Bidirectionnel) classique. La séquence vidéo est représentée par un ensemble d'images clés codées en mode I, par un ensemble de modèles 3D valides entre deux images-clés, et par une suite de positions de la caméra [24]. Une image intermédiaire entre deux images-clés est reconstruite par projection du modèle 3D texturé par une image-clé, sur le point de vue de la caméra. Les informations à transmettre (modèles 3D, positions de la caméra) sont issues de l'analyse automatique de la séquence présentée dans la section 6.1.1. La régulation de débit est réalisée d'une part par ajustement automatique de l'intervalle entre les images clés (taille des GOP) dès la phase de modélisation, et d'autre part par un codage par ondelettes des images de texture permettant une granularité fine. Cette technique de codage par ondelettes est décrite dans la section 6.2.2. Une comparaison quantitative (mesure de PSNR) et qualitative (qualité visuelle) avec des schémas de codage à bas débit normalisés (H.26L) a montré la validité de cette approche à bas débits (de l'ordre de 100 kbit/s au format CIF) et à très bas débits (16 kbit/s) pour lesquels les codeurs standards sont inopérants (voir figure 7). Noter que cette approche permet en outre de reconstruire d'autres points de vue de la scène, alliant ainsi des fonctionnalités de compression et de navigation.

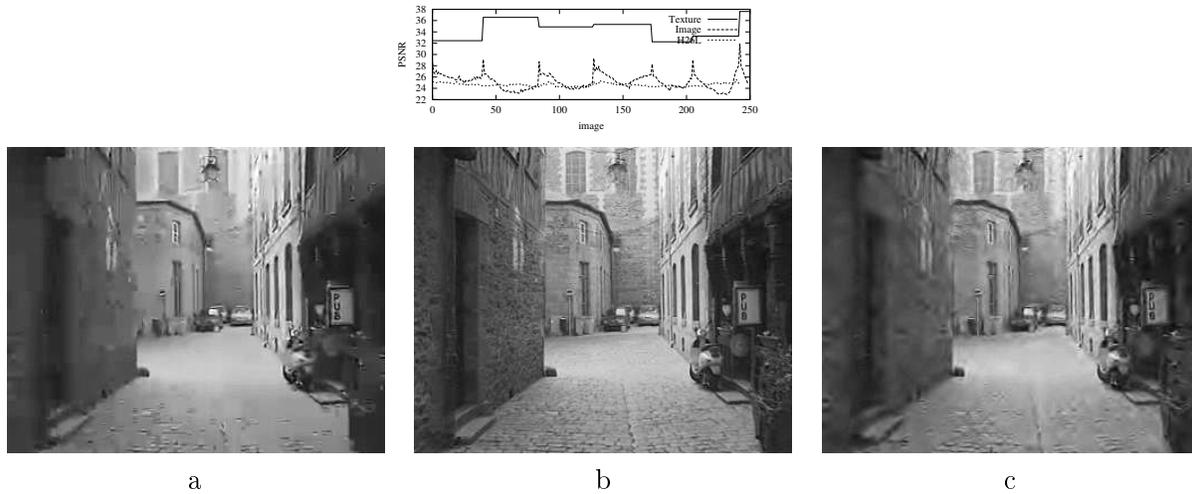


FIG. 7 – Codage par modèles 3D : Comparaison avec le codeur H26L. Pour un débit de 82kb/s, la qualité objective de notre algorithme (courbe “image”) est légèrement supérieure à celle obtenue avec H26L. La qualité de la texture avant projection (courbe “texture”) est nettement meilleure, ce qui transparait en comparant la qualité visuelle des images obtenues avec le codeur H26L (a) et notre algorithme (b). À un débit de 16 kb/s l’algorithme produit une qualité encore acceptable (c).

6.2.2 Codage scalable à grain fin

Participants : Christine Guillemot, Stéphane Pateux, Jérôme Viéron.

L’optimisation de la qualité de service d’une application de communication vidéo passe par l’adaptation dynamique du débit des signaux compressés à la bande passante du réseau de transmission. Les réseaux de transmission à qualité de service non garantie sont en effet non stationnaires et caractérisés par des variations dans le temps de la bande passante, du taux de pertes, et des délais. Dans un contexte d’application de consultation où les contenus sont pré-encodés et mis à disposition sur un serveur, cette adaptation nécessite une *représentation scalable à grain fin* des données. Il doit en effet être possible d’adapter le nombre de niveaux de scalabilité et le débit par niveau en venant sélectionner des portions du train binaire pré-encodé et stocké sur le serveur, et ceci de manière dynamique en fonction des variations des conditions de transmission.

Pour répondre à cet objectif, nous avons développé un algorithme de codage basé sur une décomposition par ondelettes spatio-temporelle de la séquence compensée en mouvement. La décomposition multi-résolution est suivie d’un codage arithmétique par plans de bits tenant compte d’une information contextuelle pour mieux exploiter les dépendances spatio-temporelles. Deux variantes de cet algorithme sont en fait disponibles, suivant la stratégie d’estimation et de compensation de mouvement utilisée. La première approche est basée sur l’outil de représentation par maillages actifs décrit dans la section 6.1.3. L’information de texture représentée par la texture dynamique, ainsi que l’information de déformation du maillage

utilisé sont codées par l'intermédiaire d'une représentation en ondelettes 3D et d'un schéma de codage de type EBCOT offrant une granularité fine de codage (granularités spatiale, temporelle, et SNR). La seconde variante repose sur une estimation et une compensation de mouvement hiérarchique basée sur une décomposition en *quadtree* contrainte en débit. Le schéma ainsi obtenu permet alors d'obtenir un niveau de performance comparable au schéma de codage H26L (TML version 6.0), tout en offrant une granularité fine de représentation.

6.2.3 Contrôle de congestion et régulation de débit

Participants : Christine Guillemot, Xavier Hénocq, Reda Hosny, Jérôme Viéron.

Pour adapter les flux vidéo à la bande passante disponible sur le réseau et qui peut fluctuer dans le temps, il est nécessaire de mettre en place dans les schémas de codage des modèles de prédiction de bande passante et de régulation du débit des signaux transmis. Un modèle de régulation de débit qui s'adapte à la fois aux modèles débit-distorsion des signaux compressés, à leur contrainte de délai maximum, et à divers modèles de prédiction de la bande passante disponible sur le réseau, a été développé [38].

Dans un contexte de transmission multi-point, un contrôle de congestion (ou de débit) réactif et précis peut exiger un nombre élevé de niveaux de scalabilité, aboutissant à des performances de compression moindres. Ce problème peut être contourné en adaptant dynamiquement le débit de chaque niveau de scalabilité sous contrainte d'un débit global constant. Dans notre approche, cette adaptation dynamique du débit est réalisée par une optimisation débit-distorsion conjointe des paramètres de quantification et des modes de codage, associée à un modèle de régulation hybride émetteur-récepteur. Le modèle de régulation hybride est associé à un mécanisme de classification (ou de *clustering*) développé par le projet PLANETE. La classification repose sur des métriques relatives aux taux de pertes ou aux délais perçus par les différents récepteurs de la liaison multi-point et permet de mettre en place une prédiction de bande passante par groupe de récepteurs. Le modèle de prédiction de bande passante repose sur des principes dits de « TCP-compatibilité », en d'autres termes sur des équations modélisant le trafic des connexions TCP, afin d'obtenir un partage équitable de la bande passante du réseau entre les données multimédia et les applications téléinformatiques. Cette approche s'inspirant des concepts des réseaux actifs permet d'adapter, de manière dynamique et avec une granularité suffisamment fine, le débit reçu par chacun des récepteurs aux conditions de transmission qu'il perçoit. L'approche permet ainsi une utilisation optimale de la bande passante pour l'ensemble des récepteurs de la session multi-point.

Les algorithmes ci-dessus sont aussi étudiés dans un environnement de différenciation de services. En particulier, une stratégie de marquage sémantique des données contenues dans un flux vidéo compressé et un mécanisme de contrôle de congestion associé à une modélisation d'un canal AF (*assured Forwarding*) de l'architecture de différenciation de service ont été développés.

6.2.4 Codage conjoint source-canal

Participants : Zhikui Chen, Christine Guillemot, Thomas Guionnet, Xavier Hénocq,

Stéphane Pateux, Gagan Rath.

Les communications sans fil sont caractérisées par des transmissions fortement bruitées avec des erreurs ponctuelles ou en rafales. L'approche classique qui consiste à considérer les deux problèmes du codage de source et de la protection (codage de canal) de manière séparée et à mettre en cascade le codeur de source, conçu pour un canal sans perte, avec des codes correcteurs d'erreurs (ex : codes convolutionnels, turbo codes, correction par anticipation, ...) s'avère sous-optimale. Malgré les bonnes performances des codes correcteurs d'erreurs, les contraintes d'un système réel (complexité, délai de décodage,...) conduisent à des taux d'erreurs résiduels non négligeables. Il est donc apparu nécessaire d'optimiser les deux opérations de codage (source et canal) conjointement.

Les opérateurs de compression ont été conçus à ce jour de manière d'une part à décorréler le signal (transformations) et d'autre part à optimiser les performances débit-distorsion de quantification. Le codage de canal a pour objectif l'ajout de redondance dans le train binaire transmis sur le réseau afin de pouvoir corriger les erreurs de transmission. Dans le modèle utilisant des codes correcteurs associés aux flux vidéo transmis, les performances du système de communication peuvent aussi être améliorées par la mise en place de techniques de décodage conjoint source-canal où une connaissance *a priori* des statistiques de la source est exploitée au niveau du décodage de canal (voir section 6.2.5). Il est aussi possible d'introduire de la redondance directement dans le flux compressé, en concevant des codeurs de source qui puissent s'adapter à des caractéristiques et modèles du canal variant dans le temps.

Dans ce contexte, nos travaux ont porté d'une part sur la conception et l'étude de transformations redondantes permettant d'introduire une quantité « contrôlée » de corrélation ou d'informations redondantes, et d'autre part sur le codage par descriptions multiples basé sur une quantification redondante. Noter que le codage par descriptions multiples est une extension du codage avec critère de fidélité, suivant la théorie débit distorsion, au cas multi-canal. L'interprétation de ces transformations redondantes sous forme de codes de canal permet d'appliquer certains outils de décodage de canal (comme par exemple les algorithmes de décodage basés syndrome ou des estimateurs Bayésiens) pour exploiter la redondance et corriger les erreurs (ou pertes) dues à la transmission. Nous avons développé une approche de codage conjoint basé sur une transformation redondante à la fois dans un contexte de pertes de paquets (Internet filaire) et de perturbations comprenant à la fois des pertes et du bruit impulsif (Internet sans fil). Une analyse des performances théoriques du code conjoint source-canal ainsi défini a permis de définir une stratégie de paquetisation d'une part et un algorithme de décodage itératif par syndrome d'autre part permettant d'améliorer significativement la qualité de la chaîne de transmission de bout-en-bout en présence de pertes consécutives [34]. Le couplage du décodage par syndrome avec des techniques de projection sur des ensembles convexes a aussi permis de réduire l'impact du bruit de quantification sur la fiabilité de l'estimation des symboles perturbés par le bruit de transmission.

Une approche complémentaire basée sur une quantification redondante a aussi été étudiée. Cette technique a été incorporée dans un codeur d'images fixes basé sur l'algorithme normalisé par JPEG2000. L'approche permet de répondre en outre à des objectifs supplémentaires de transmission progressive des signaux à des fins d'adaptation du débit aux caractéristiques variables du canal de transmission [26].

6.2.5 Décodage conjoint source-canal

Participants : Francisco Garcia-Ugalde, Christine Guillemot, Thomas Guionnet, Pierre Siohan.

Les objectifs de forte compression conduisent à l'utilisation de codes statistiques à longueurs variables. Ces codes sont très sensibles au bruit, une erreur pouvant entraîner une désynchronisation du décodeur. Une modélisation sous forme d'automates stochastiques des codeurs à longueurs variables a permis de construire un algorithme de décodage robuste de codes à longueurs variables (codes d'Huffman) et de codes à longueurs variables réversibles (RVLC) [15]. Ce travail a été mené en collaboration avec le projet SIGMA2. Le décodage basé sur une estimation Bayésienne permet de réduire fortement les taux d'erreurs symboles et d'accroître la capacité de resynchronisation du décodeur mesurée à l'aide de la distance de Levenshtein. La modélisation sous forme de chaînes de Markov des dépendances entre les symboles de source d'une part, et entre les bits de la séquence codée par le code à longueur variable d'autre part, fournit un cadre naturel à l'introduction de petits mots de synchronisation permettant une synchronisation « douce » du décodeur. Par rapport à des codes auto-synchronisants, ces mots de synchronisation ne doivent pas vérifier les conditions de préfixe des codes à longueurs variables et présentent donc une redondance moindre [28]. L'application du principe turbo (ou de propagation de croyance) entre le décodeur de codes à longueurs variables et le décodeur de canal utilisés en alternance selon une structure série a permis d'atteindre des taux d'erreurs symboles très faibles pour des taux d'erreurs bits au niveau du canal de l'ordre de 5×10^{-2} . Les algorithmes ont d'abord été testés sur des sources de Gauss-Markov dans un contexte de description unique de la source et dans un contexte de descriptions multiples supposant une transmission sur des canaux multiples non corrélés. Nous avons aussi démarré l'adaptation et l'intégration de ces outils dans un décodeur vidéo basé sur la norme H.263+ afin de valider les approches sur des signaux réels. L'extension de ce formalisme et de ces outils aux codes arithmétiques qui tendent à occuper une place de plus en plus importante dans les schémas de compression émergents est en cours.

6.3 Tatouage d'images

Participants : Gaëtan Le Guelvouit, Stéphane Pateux, Christine Guillemot.

Mots clés : tatouage, signature, copyright, étalement de spectre, codage canal.

Résumé : *Notre approche en tatouage cherche à exploiter d'une part, et à mieux formaliser, voire quantifier, d'autre part, le parallèle existant entre les techniques de tatouage d'images et de transmission d'informations à travers un canal bruité. Le message à insérer peut être vu comme un message à transmettre sur un canal bruité, à savoir l'image, qui peut subir des attaques de formes diverses (compression, filtrage, transformations géométriques, etc). Ce parallèle conduit naturellement à considérer les codes correcteurs comme des outils de protection de la marque, de manière à permettre in fine une détection et une extraction robuste. L'apport des codes correcteurs par rapport à la robustesse du tatouage face à diverses attaques*

(compression JPEG, retaille, filtrage,...) a été étudié et quantifié.

Le problème du tatouage d'image peut être formalisé comme un problème de communication sur un canal bruité. L'objectif est en effet d'insérer le plus grand nombre possible de bits d'information dans le signal sous contrainte d'une distorsion donnée entre le signal original et le signal tatoué, tout en permettant une estimation fiable des bits insérés en présence d'une distorsion d'attaque donnée. Le problème peut alors être formulé comme un problème de maximisation-minimisation conjointe de la capacité (ou du nombre de bits insérés) sous contraintes respectives de distorsions d'insertion et d'attaque. La résolution de ce problème peut aussi être vu comme un exemple d'application de la théorie du jeu : le jeu de l'attaquant consiste à trouver des attaques minimisant la capacité sous contraintes de distorsions d'attaques données et celui du défenseur consiste à chercher des solutions d'insertion au contraire maximisant cette capacité. Ce formalisme nous a conduit à caractériser l'attaque optimale associée au canal équivalent à la chaîne de tatouage, dans un contexte d'insertion par étalement de spectre et d'utilisation d'une métrique de distorsion perceptuelle. Nous avons alors analysé les limites de performances théoriques en terme de capacité d'insertion en fonction des distorsions d'insertion et d'attaques (distorsion prenant en compte une métrique psychovisuelle). Un schéma d'insertion et d'extraction permettant d'atteindre ces limites a également été conçu. Le schéma a été testé et validé en présence d'une large gamme d'attaques non géométriques, intentionnelles ou non. Nous avons également montré l'intérêt de l'utilisation de codes correcteurs, et notamment étudié l'influence du rendement des codes utilisés sur les performances globales. L'extension de l'approche actuelle aux attaques géométriques est en cours.

6.4 Résultats relatifs à l'action TexMex en cours de montage

Participants : Laurent Amsaleg, Sid-Ahmed Berrani.

Mots clés : Indexation, bases de données.

En collaboration avec Patrick Gros du projet VISTA, nous avons poursuivi notre étude des problèmes posés par l'indexation par le contenu et la recherche dans de grandes bases de documents multimédias. Ces problèmes relèvent à la fois des techniques issues du domaine de l'image (calculs de descripteurs, reconnaissance) et des bases de données (gestion de gros volumes d'information sur disques).

Nous avons d'abord réalisé une étude comparative des meilleures techniques d'indexation issues des bases de données mais utilisées dans un contexte différent de celui pour lequel elles ont été conçues, à savoir celui de la reconnaissance d'image de grain fin, techniques images modernes et très performantes (et non pas de la reconnaissance gros grain). Nous montrons dans [11, 21] que ces schémas d'indexation sont inutilisables en pratique dans ce contexte, et que leurs performances sont parfois moins bonnes que celles d'une recherche exhaustive sans index. Une recherche exhaustive n'est toutefois pas viable lorsqu'il s'agit de traiter plusieurs Tera-octets.

Aussi, nous avons commencé à explorer ces différentes pistes [21]. Brièvement, nous essayons de réduire par des techniques de *clustering* le nombre de points à examiner, d'utiliser le plus souvent possible des approximations des points restants plutôt que les données originales, de

coder ces points de façon à ce que l'espace devienne, par contraction, monodimensionnel, et de procéder à des recherches couplées à une évaluation du pourcentage de confiance dans la qualité et la stabilité du résultat afin d'interrompre l'exploitation dès qu'il y a convergence. Cette approche se démarque des propositions faites par ailleurs parce qu'elle mélange plusieurs techniques disjointes, qu'elle utilise pleinement les caractéristiques particulières d'une famille de descripteurs (descripteurs locaux) et qu'elle aborde frontalement le problème de la dégradation de la qualité du résultat.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Contrats industriels bi-latéraux

7.1.1 CTI-France Télécom R&D

Participants : Christine Guillemot, Xavier Hénocq.

- Numéro de la convention : 1 98 C 391 00 31324 01 2.
- Intitulé : *Transmission multi-point de données multimédia sur réseaux hétérogènes.*
- Activité de recherche concernée : § 6.2.4 et § 6.2.3.
- Partenaire : Inria, France Télécom R&D.
- Financement : CTI-France Télécom R&D.
- Période : 36 mois à partir de septembre 1998.

Notre contribution porte sur les thèmes de recherche suivants :

- l'étude de techniques de codage vidéo robustes aux pertes de paquets ;
- l'étude de modes de représentation scalable permettant un bon compromis entre décorrélation et robustesse aux pertes de paquets ;
- l'étude de mécanismes de codage conjoint source-canal, permettant un partage de bande passante optimal entre le débit utile et le débit de redondance ;
- l'étude de mécanismes de masquage de pertes adaptés à l'Internet ;
- l'étude et le développement de mécanismes modifiés et adaptés à un Internet multi-service.

7.1.2 Convention Celar

Participants : Thierry Fauconnier, Christine Guillemot, Gagan Rath.

- Numéro de la convention : 1 99 C 459 00 31324 061 et 1 00 C 0302 00 31324 06 1.
- Intitulé : *Liaison vidéo numérique en robotique mobile.*
- Activité de recherche concernée : § 6.2.5.
- Partenaire : Inria, Celar.
- Financement : Celar.
- Période pour la phase 1 : 6 mois à partir de septembre 1999.
- Période pour la phase 2 : 18 mois à partir de septembre 2000.

Le problème posé consiste à transmettre un signal vidéo sur un réseau sans fil à partir d'un équipement d'acquisition mobile, pour une application de téléguidage d'un robot terrestre

mobile. Le réseau support est ainsi caractérisé par une faible bande passante et une qualité de service non garantie qui peut en outre varier dans le temps. Afin de permettre une utilisation optimale de la bande passante disponible à un instant donné et variant dans le temps, la source vidéo émettrice doit prendre en compte les caractéristiques du canal. Le codeur doit donc émettre un débit relativement élevé mais robuste aux erreurs de transmission. Nous nous intéressons dans ce contexte au problème du décodage conjoint source-canal, et au codage par descriptions multiples.

7.1.3 Convention Thomson-TCO

Participant : Henri Nicolas.

- Numéro de la convention : 1 98 C 307 00 31324 012.
- Intitulé : *Création interactive d'objets mosaïques 2D1/2 et 3D*.
- Activité de recherche concernée : Thèse d'Agnès De Simon, § 5.
- Partenaires : Inria, Thomson-TCO.
- Financement : Convention Cifre Thomson-TCO.
- Période : 36 mois à partir de mai 1998.

Cette convention Cifre avec Thomson Optronique concerne l'encadrement de la thèse d'Agnès De Simon sur le thème : « création interactive d'objets mosaïques 2D1/2 et 3D ». L'objectif de cette étude est de développer de nouveaux outils permettant la reconstitution interactive de scènes 3D à partir d'outils de calibration interactive et d'algorithmes de création d'images mosaïques.

7.1.4 Convention Thomson Multimédia

Participants : Luce Morin, Philippe Robert.

- Activité de recherche concernée : Thèse de Yannick Nicolas, § 6.1.1.
- Partenaires : Inria, Thomson Multimédia.
- Financement : Convention Cifre Thomson Multimédia.
- Période : 36 mois à partir de octobre 1998.

L'objectif de cette convention Cifre est d'étudier une chaîne de navigation à distance dans des modèles de scènes 3D réelles. Plus précisément, il s'agit d'étudier et de développer les éléments d'une chaîne de modélisation, de transmission et de rendu pour la navigation virtuelle dans des scènes réelles complexes, dans un contexte de diffusion et d'applications client-serveur. Les systèmes actuels permettent la navigation dans des conditions limitées quant à la qualité des images et au réalisme de la scène, et à la flexibilité de navigation, elle-même restreinte par le type de mouvement de la caméra virtuelle (panoramique, zooms, translation, etc.).

7.2 Contrats nationaux

7.2.1 RNRT-V2net

Participants : Franck Galpin, Luce Morin.

- Numéro de la convention : 2 00 C 0339.
- Intitulé : Projet RNRT-V2NET, *Visite virtuelle en réseau*.
- Activité de recherche concernée : § 6.1.1.
- Partenaires : Thomson Multimédia, Inria, France Télécom R&D.
- Financement : RNRT.
- Période : 36 mois à partir de janvier 2000.

L'objectif du projet V2NET est de développer et intégrer de nouveaux outils pour une navigation optimale dans des mondes 3D réels et virtuels dans un contexte d'applications client-serveur. L'optimalité est définie ici comme la recherche d'un meilleur compromis qualité d'image/fluidité de navigation. Les études portent plus précisément sur la modélisation et la représentation de scènes 3D complexes réelles et virtuelles, la structuration de la scène et les protocoles applicatifs, la visualisation de scènes 3D complexes réelles et virtuelles, et la mise-en-place d'une plate-forme d'expérimentation permettant de valider les approches proposées.

7.2.2 RNRT-Osiam

Participant : Stéphane Pateux.

- Numéro de la convention : 298C5620031324011.
- Intitulé : Projet RNRT-Osiam, *Outils de segmentation spatio-temporelle pour MPEG-4/MPEG-7*.
- Activité de recherche concernée : § 6.1.2.
- Partenaires : Philips-Lep, Polytech. Nantes, Irisa/Inria-Rennes, I3S, Timc, Creatis, France Télécom R&D.
- Financement : RNRT.
- Période : 36 mois à partir de décembre 1998.

Le projet exploratoire RNRT-Osiam a pour objectif la spécification d'outils de segmentation spatio-temporelle pour MPEG-4/MPEG-7. Diverses approches complémentaires de segmentation hiérarchique et de suivi d'objets vidéo sont mises en oeuvre, comparées, intégrées au sein d'une même plate-forme logicielle. Les domaines applicatifs visés sont la vidéosurveillance, la télémédecine et la composition d'objets vidéo pour des applications de post-production, et d'indexation de contenus vidéo.

7.2.3 RNRT-Cohrainte

Participant : Stéphane Pateux.

- Numéro de la convention : 299C3220031324011.
- Intitulé : Projet RNRT-Cohrainte, *Codage hiérarchique et robuste de sources audiovisuelles et application à l'Internet*.
- Activité de recherche concernée : § 6.1.3.
- Partenaires : France Télécom R&D, Philips-Lep, Irisa/Inria-Rennes, ENST-Paris.
- Financement : RNRT.
- Période : 36 mois à partir de mars 1999.

Le projet exploratoire Cohrainte s'intéresse à la spécification de sources scalables à granularité fine pour la représentation de flux de données audio-visuelles. L'application visée étant le transport de ces flux sur réseaux IP, une attention particulière est portée à l'allocation conjointe, et négociée dynamiquement, des ressources aux flux audio et vidéo.

7.2.4 RNRT-Cosocati

Participant : Pierre Siohan.

- Intitulé : *COdage conjoint SOurce-canal pour la Transmission d'Images*.
- Activité de recherche concernée : § 6.2.5
- Partenaires : France Télécom R&D (coordonnateur), Alcatel Space Industries, CNES, GET, I3S.
- Financement : RNRT.
- Période : 36 mois à partir de mars 2000.

Avec l'arrivée en septembre de P. Siohan, en position de détachement de France Télécom R&D, l'Irisa-Inria Rennes va apparaître comme un nouveau partenaire de ce projet à caractère exploratoire. Un appel à candidature pour un post-doctorant, annoncé dans la rubrique emplois du site de l'Irisa (<http://www.irisa.fr/accueil/emplois/temics-2.htm>) et de l'INRIA, définit plus en détails ce que sera notre contribution à ce projet. Plus généralement, Cosocati concerne l'application de techniques de codage et décodage conjoint source-canal pour de la transmission d'images dans le cas de systèmes de télécommunications utilisant des canaux de type radio-mobile, et dans celui de systèmes d'observation spatiale par satellite. On espère ainsi tirer parti d'avantages déjà reconnus du codage conjoint : un gain de complexité de calcul ou de matériel pour un niveau de performances donné, une zone de fonctionnement nominale étendue par rapport aux systèmes de communications classiques, ou dit d'une autre façon dans le cas de la diffusion, une réduction de la puissance émise pour une même zone de couverture. Plusieurs techniques très prometteuses de codage combiné source-canal ont déjà été étudiées et développées par les universitaires participant à ce consortium, l'objectif visé est de les expérimenter dans le cadre d'applications réelles liées à la transmission d'images.

7.2.5 RNRT-Visi

Participants : Bilal ElAli, Christine Guillemot, Jérôme Viéron.

- Numéro de la convention : 2 99 A 461 00 000MC 011.
- Intitulé : *vidéo interactive sur Internet*.
- Activité de recherche concernée : § 6.2.3.
- Partenaires : Thomson Multimédia, Inria, Edixia, Eurecom, France Télécom R&D.
- Financement : RNRT.
- Période : 26 mois à partir de septembre 1999.

Le projet a pour objectif de fournir des solutions au problème de la gestion de la qualité de service dans des applications de transmission vidéo sur l'Internet. Il étudie différentes techniques de représentation scalable à grain fin des sources avec des mécanismes de régulation de débit appropriés, de contrôle de congestion et le contrôle d'erreur. Il étudie également les

concepts architecturaux et systèmes liés à la répartition des sources entre plusieurs serveurs et les techniques de cache vidéo. L'objectif est en outre de développer des solutions aussi pérennes que possible, à savoir qui puissent d'une part s'appliquer à l'Internet *best-effort* d'aujourd'hui, mais être également compatibles avec une évolution – à l'étude au sein de l'IETF (*Internet Engineering Task Force*) – du réseau vers l'intégration et la différenciation de services. En 2001, TEMICS a, en étroite collaboration avec le projet PLANETE de l'INRIA Sophia-Antipolis, développé un algorithme de contrôle de congestion et de débit en multi-point. Cet algorithme est couplé avec un mécanisme d'agrégation d'informations statistiques sur l'état du réseau développé par le projet PLANETE et basé sur un algorithme de classification de récepteurs suivant des métriques faisant intervenir le taux de pertes, la bande passante disponible, le débit TCP-compatible, etc ... Ces algorithmes ont été intégrés à la plate-forme de communication développée au sein du projet TEMICS (voir section 5.1).

7.3 Contrats européens

7.3.1 IST-Openise

Participants : Christine Guillemot, Reda Hosny, Gagan Rath.

- Numéro de la convention : 1 99 A 1800 00 00MPR.
- Intitulé : Convention IST-OPENISE, *Open Platform for Enhanced Interactive Services*.
- Activité de recherche concernée : § 6.2.3.
- Partenaires : Italtel, France Télécom, DT-Berkom, FHW, Inria, Optibase, Orkit, Siemens, VideoTime.
- Période : 24 mois à partir de janvier 2000.
- Financement : CEE.

L'objectif du projet est le développement d'une plate-forme de communication sur Internet incorporant des applications interactives telles que mondes virtuels partagés, télévision interactive... Les nouvelles technologies développées dans le projet concernent le codage robuste, le contrôle de transmission des contenus dans un contexte de différenciation de services, les outils de modélisation et de partitionnement de mondes virtuels permettant de faciliter la navigation au travers d'un réseau de communication et qui permettent également d'adapter les contenus transmis à la navigation de l'utilisateur ainsi qu'à la bande passante disponible sur le réseau. Dans le cadre de ce projet, TEMICS se concentre sur les questions de différenciation de services et a dans ce contexte développé un algorithme de régulation de débit des flux vidéo adapté à un modèle de canal associé à la classe de service AF (*Assured Forwarding*).

7.3.2 IST-Song

Participants : Bilal ElAli, Christine Guillemot.

- Numéro de la convention : 1 99 A 1821 00 00MP R00 5.
- Intitulé : IST-SONG, *Portals of Next Generation*.
- Activité de recherche concernée : § 5.1.
- Partenaires : France Télécom, DT-Berkom, Blaxxun, Cselt, Enst-br, Epfl, GMD, Inria, RUS, Telenor, Univ. Genève.

- Période : 34 mois à partir de janvier 2000.
- Financement : CEE.

L'objectif du projet est de développer de nouvelles technologies d'interfaces hommes-machines (avatars, agents intelligents, etc.), de nouveaux algorithmes de représentation et de communication de contenus audio-visuels et de mondes virtuels sur Internet et sur des liens sans fils, pour des « *portails* » de seconde génération. En collaboration étroite avec le CSELT, France Télécom et RUS, TEMICS a contribué à la définition de stratégies de transport de données multimédia sur Internet, à l'étude d'algorithmes de contrôle de congestion et de codage conjoint source canal sur l'Internet « *best-effort* » et dans un cadre de différenciation de services.

8 Actions régionales, nationales et internationales

8.1 Réseaux et groupes de travail internationaux bilatéraux

Le projet Temics participe également à plusieurs collaborations bi-latérales :

- programme France-Maroc (participants : C. Labit, L. Morin, S. Pateux) : cette action intégrée franco-marocaine (responsable français : L. Morin) est menée en collaboration avec l'université de Rabat, Maroc (Professeur D. Aboutajdine, professeur assistant de l'INPT-Rabat A. Tamtaoui) ; Participation de S. Pateux au colloque France-Maroc, avril 2001 ;
- programme France-Mexique (participants : L. Morin, H. Nicolas, S. Pateux, C. Guillemot) : cette collaboration est menée avec l'Unam-Mexico (professeur Garcia-Ugalde) dans le cadre d'une convention CNRS-Conacyt ;
- participation au colloque France-Taiwan (participants : C. Guillemot), Taiwan, Mars 2001.
- participation au colloque INRIA-HKU, avec en charge l'organisation d'une session sur le thème du multimédia, dans le cadre de la signature d'un accord de collaboration entre l'INRIA et HKU (Hong-Kong University) (participant : C. Guillemot), Hong-Kong, Avril 2001 ;
- collaboration avec R. Ansari de l'Université d'Illinois à Chicago (UIC) et avec P. Duhamel du CNRS sur le thème du codage conjoint source-canal ; il s'agit d'une ARC IRISA dans le cadre de laquelle nous accueillons S. Marinkovic sur un statut postdoctoral ;
- collaboration avec l'Université Catholique de Louvain (UCL) ; il s'agit d'une collaboration avec B. Macq sur le thème du tatouage, dans le cadre de laquelle nous accueillons T. Furon sur un statut postdoctoral.

8.2 Actions nationales

8.2.1 ACI Jeunes chercheurs

Participants : Laurent Amsaleg, Sid-Ahmed Berrani.

- Intitulé : *Techniques d'exploitation des données multimédias*.
- Activité de recherche concernée : § 6.4.

- Partenaires : VISTA
- Financement : ministère de la recherche.
- Période : 36 mois à partir de juillet 2001.

Ce programme du ministère de la recherche a pour but d'aider et d'accompagner la création de nouvelles équipes par de jeunes chercheurs. Le projet qui a été accepté vise à monter une nouvelle équipe sur le thème de l'indexation, de la recherche et plus largement de l'exploitation des grandes bases de documents multimédias. Cette équipe sera effectivement créée sous forme d'action INRIA au 1^{ier} janvier 2002.

8.2.2 Action CNRS JemSTIC

Participants : Laurent Amsaleg, Sid-Ahmed Berrani.

- Numéro de la convention : STIC 01 N80/0490
- Intitulé : Techniques d'exploitation des données multimédia
- Activité de recherche concernée : § 6.4.
- Partenaires : VISTA
- Financement : CNRS.
- Période : 12 mois à partir d'octobre 2001.

Cette action, intitulée « Jeune Equipe, Mobilité et Jeune Chercheur », et émanant du CNRS, a pour but de favoriser l'émergence de jeunes chercheurs, de nouvelles équipes, de nouvelles thématiques et de favoriser les mobilités allant dans ce sens. Le soutien concerne le projet de nouvelle équipe sur le thème de l'indexation, de la recherche et plus largement de l'exploitation des grandes bases de documents multimédias.

9 Diffusion de résultats

9.1 Animation de la communauté scientifique

- L. Amsaleg est co-animateur du groupe de travail CIDRE ;
- L. Amsaleg est conférencier invité aux journées Entrepôts de données et décisionnel, novembre 2001, Ifsic ;
- L. Amsaleg est conférencier invité aux 9èmes Rencontres INRIA-Industrie ayant pour thème la « Recherche d'information dans les documents électroniques » ;
- C. Guillemot est éditeur associé du journal IEEE transactions on Image Processing ;
- C. Guillemot est membre du comité de programme de la conférence Spie-VCIP 2001, IEEE-ICIP 2001 ;
- C. Guillemot est membre du comité technique IMDSP (Image and MultiDimensional Signal Processing Technical Committee) ;
- C. Guillemot est membre de la commission 2 du RNRT ;
- L. Morin, H. Nicolas et S. Pateux ont participé activement (par le biais de présentations) aux réunions du GDR Isis ;
- H. Nicolas est membre nommé (suppléant) de la commission de spécialiste en 27ième section ;

- S. Pateux est coordinateur scientifique au niveau de l'IRISA du CNRT (Centre National de Recherche Technologique) TIM-Bretagne ; Organisation de la journée thématique 'indexation multimédia', Irisa, 11 mai 2001.

9.2 Enseignement

- Diic-LSI, Ifsic, université de Rennes 1 (H. Nicolas : codage) ;
- Diic-INC, Ifsic, université de Rennes 1 (L. Morin, C. Labit, H. Nicolas, F. Garcia Ugalde, C. Guillemot : traitement d'images, vision 3D, mouvement, codage, compression d'images, cryptographie, communication) ;
- DEA Stir, option Image, université de Rennes 1 (C. Labit, H. Nicolas : codage) ;
- DEA Informatique, université de Rennes 1 (C. Guillemot, S. Pateux : Compression Vidéo et communication) ;
- DEA TIS, université de Cergy-Ensea (C. Guillemot : codage) ;
- Enic, Villeneuve-d'Ascq (C. Guillemot : Vidéo sur IP) ;
- Ensar Rennes, option informatique (L. Morin : fondements du traitement d'image, morphologie mathématique) ;
- ENSTBR (C. Guillemot : Les normes de codage vidéo de MPEG-1 à MPEG-4) ;
- DEA ÉLECTRONIQUE, INSA, université de Rennes 1 (P. Siohan : Télévision numérique) ;
- IUP-3, Ifsic, Université de Rennes 1 : Laurent Amsaleg – trois aspects des bases de données : parallélismes, images et cartes à puce ;
- DESS Mitic, Ifsic, Université de Rennes 1 : Laurent Amsaleg – Le document numérique et l'indexation multimédia : aspects généraux, images, texte et support des bases de données.

10 Bibliographie

Ouvrages et articles de référence de l'équipe

- [1] C. GUILLEMOT, R. ANSARI, « Layered Coding Schemes for Video Transmission on ATM Networks », *Journal of Visual Communication and Image Representation* 5, 1, March 1994, p. 62–74.
- [2] C. GUILLEMOT, T. TURLETTI, *Systèmes multimédia communicants*, Éditions Hermès, 2000, ch. Algorithmes de codage et de contrôle de transmission sur réseaux hétérogènes, p. 25–82.
- [3] J. P. LEDUC, J. M. ODOBEZ, C. LABIT, « Adaptive motion-compensated wavelet filtering for image sequence coding », *IEEE Transactions on Image Processing* 6, 6, June 1997, p. 862–879.
- [4] R. MOHR, L. MORIN, E. GROSSO, « Relative positioning with uncalibrated cameras », in : *Geometric Invariance in Computer Vision*, Editors : A. Zisserman J.L. Mundy, MIT Press, 1992, p. 440–460.
- [5] F. MOREAU DE SAINT-MARTIN, P. SIOHAN, A. COHEN, « “Biorthogonal filterbanks and energy preservation in image compression” », *IEEE Transactions on Image Processing* 8, 2, février 1999, p. 168–178.
- [6] H. NICOLAS, C. LABIT, « Motion and illumination variation estimation using a hierarchy of models : application to image sequence coding », *Journal of Visual Communication and Image Representation* 6, 4, December 1995, p. 303–316.

- [7] P. RAULT, C. GUILLEMOT, « Symmetric delay factorization : genralized framework for paraunitary filter banks », *IEEE Trans. on signal processing* 47, 12, Dec. 1999, p. 3315–3325.
- [8] P. SIOHAN, C. ROCHE, « Cosine modulated filterbanks based on Extended Gaussian Functions », *IEEE Transactions on Signal Processing* 48, 11, novembre 2000, p. 3052–3061.
- [9] G. TZIRITAS, C. LABIT, *Motion Analysis for Image Sequence Coding*, Elsevier Science, Book in Series “Advances in Image Communication”, 390 pages, 1994.

Thèses et habilitations à diriger des recherches

- [10] H. NICOLAS, *Contribution à la création et à la manipulation des objets vidéo*, thèse de doctorat, Fév. 2001.

Articles et chapitres de livre

- [11] L. AMSALEG, P. GROS, « Content-based Retrieval Using Local Descriptors : Problems and Issues from a Database Perspective », *Pattern Analysis and Applications* 4, 2/3, 2001, p. 108–124.
- [12] J. BENOIS, H. NICOLAS, « A new method for region-based depth ordering in a video sequence : application to frame interpolation », *Accepté pour publication dans Journal of Visual Communication and Image Representation*.
- [13] J. BENOIS-PINEAU, S. PATEUX, D. BARBA, *Compression et codage des images et des vidéos*, Hermès, 2001, ch. 5 : Modèles de représentation d’images vidéo et compression, p. 133–177.
- [14] C. GUILLEMOT, S. PATEUX, *Compression et codage des images et des vidéos*, Hermès, 2001, ch. 1 : Eléments de théorie de l’information et de communication vidéo, p. 21–43.
- [15] A. GUYADER, E. FABRE, C. GUILLEMOT, M. ROBERT, « Joint Source-Channel Turbo Decoding of Entropy-Coded Sources », *IEEE Journal on Selected Areas in Communications, special issue on the turbo principle : from theory to practice* 19, 9, sept. 2001, p. 1680–1696.
- [16] H. NICOLAS, C. GUILLEMOT, *Compression et codage des images et des vidéos*, Hermès, 2001, ch. 6 : Normes de compression vidéo, p. 178–200.
- [17] H. NICOLAS, S. PATEUX, D. LEGUEN, « Minumum Description Length Criterion and Segmentation Map Coding for Region-based Video Compression », *IEEE Transactions on Circuits and Sstems for Video Technology* 11, 2, fév. 2001, p. 184–198.
- [18] H. NICOLAS, « Création et manipulation des objets vidéo », *Techniques et sciences informatiques* 20, 8, oct. 2001.
- [19] H. NICOLAS, « New methods for Dynamic Mosaicking », *IEEE Transactions on Image Processing* 10, 8, août 2001, p. 1239–1251.
- [20] P. RAULT, C. GUILLEMOT, « Indexing algorithms for Z_n , A_n , D_n and D_n^{++} lattice vector quantizers », *IEEE Trans. on Multimedia* 3, 4, dec. 2001.

Communications à des congrès, colloques, etc.

- [21] L. AMSALEG, P. GROS, S.-A. BERRANI, « A Robust Technique to Recognize Objects in Images, and the DB Problems it Raises », *in : Workshop on Multimedia Information Systems*, Capri, Italy, nov. 2001.
- [22] Z. CHEN, C. GUILLEMOT, R. ANSARI, « Multiple description coding based on multiwalet transforms », *in : Proceedings of the Picture Coding Symposium, PCS’001*, mai 2001.

- [23] E. FABRE, A. GUYADER, C. GUILLEMOT, « Joint source and channel turbo decoding of VLC-coded Markov sources », in : *Proceedings of the IEEE Int. Conference on ASSP (ICASSP'2001)*, avril 2001.
- [24] F. GALPIN, L. MORIN, « Computed 3D Models for Very Low Bitrate Video Coding », in : *Proceedings of the SPIE Conference on Visual Communications and Image Processing, San Jose, USA*, janv. 2001.
- [25] F. GALPIN, L. MORIN, « Reconstruction 3D euclidienne par localisation à partir d'un calibrage inexact », in : *Congrès francophone de Vision par Ordinateur, ORASIS 2001, Cahors, France*, p. 57–66, juin 2001.
- [26] T. GUIONNET, C. GUILLEMOT, S. PATEUX, « Embedded multiple description coding for progressive image transmission over unreliable channels », in : *Proceedings of the IEEE Int. Conference on Image Processing, ICIP 2001*, oct. 2001.
- [27] A. GUYADER, E. FABRE, C. GUILLEMOT, « Joint source and channel turbo decoding of VLC-coded Markov sources », in : *Proceedings de la conférence GRETSI 2001*, sept. 2001.
- [28] A. GUYADER, E. FABRE, C. GUILLEMOT, « Robust decoding of VLC encoded Markov sources », in : *Proceedings de la conférence GRETSI 2001*, sept. 2001.
- [29] X. HÉNOCQ, G. RATH, C. GUILLEMOT, « Correction de pertes par expansion de signal sur bases de fonctions redondantes », in : *Proceedings of CORESA 2001*, nov. 2001.
- [30] H. NICOLAS, J. FIGUE, « Semi-automatic Object-based Video Manipulations for Post-production and Video Editing Applications », in : *Proceedings of SCI, pp. 310-315, Vol. IX, Orlando, USA*, juil. 2001.
- [31] S. PATEUX, G. MARQUANT, D. CHAVIRA-MARTINEZ, « Object mosaicking via meshes and crack-lines technique. Application to low bit-rate video coding », in : *Proceedings of Picture Coding Symposium 2001*, p. 417–420, Séoul, Korea, avril 2001.
- [32] J. PINEL, H. NICOLAS, « Estimation of 2D illuminant direction and shadow segmentation in natural video sequences », in : *Proceedings of VLBV, pp. 197-202, Athens, Greece*, oct. 2001.
- [33] J. PINEL, H. NICOLAS, « Méthodes d'analyse/synthèse des ombres dans une séquence d'images naturelles », in : *Proceedings of CORESA, Dijon, France*, nov. 2001.
- [34] G. RATH, X. HÉNOCQ, C. GUILLEMOT, « Application of DFT codes for robustness to erasures », in : *Proceedings of the Int. Globecom Conference, 2001*, nov. 2001.
- [35] M. RZIZA, D. ABOUTAJDINE, L. MORIN, A. TAMTAOUI, « New disparity map estimation using higher order statistics », in : *Proceedings of the IEEE Conference on Acoustic, Speech, and Signal Processing, Salt Lake City, USA, III*, mai 2001.
- [36] C. SICLET, P. SIOHAN, D. PINCHON, « Analyse et conception des systèmes multiporteuses BFDM suréchantillonnés », in : *Proceedings de la conférence GRETSI 2001*, sept. 2001.
- [37] J. VIÉRON, C. GUILLEMOT, « Contrôle de débit TCP-compatible pour la vidéo sur l'Internet », in : *Proceedings of CORESA 2001*, nov. 2001.
- [38] J. VIÉRON, C. GUILLEMOT, « Source adaptive TCP-compatible rate control for video over the Internet », in : *Proceedings of the IEEE Int. Conference on Image Processing, ICIP 2001*, oct. 2001.