

Projet TEMICS

Traitement, Modélisation d'Images et CommunicationS

Rennes

THÈME 3B



*R*apport
d'Activité

1999

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	4
2.1	Objectifs et Méthodologie	4
2.2	Contexte international	5
2.3	Axes de recherche	6
3	Fondements scientifiques	7
3.1	Modélisation et segmentation d'objets vidéo	7
3.1.1	Segmentation et suivi temporel d'objets Vidéo	7
3.1.2	Modélisation 3D par analyse de séquences vidéo, compression et reconstruction de vues	9
3.2	Analyse, manipulation d'objets vidéo et codage de l'information associée	11
3.2.1	Objectifs	11
3.2.2	Approches	12
3.2.3	Fondements scientifiques	13
3.3	Couplage source-réseau	13
3.3.1	Codage conjoint pour réseaux hétérogènes en se focalisant sur les réseaux IP	14
3.3.2	Codage conjoint source-canal pour communications sans fils	15
3.4	Tatouage d'images et de contenus vidéo	16
3.5	Thèmes transversaux: scalabilité et sélectivité	17
4	Domaines d'applications	17
4.1	Panorama	17
4.2	Compression d'images fixes et animées	18
4.3	Communication multimédia	18
4.4	Production et post-production vidéo	19
5	Logiciels	19
5.1	Plate-forme de communication multimédia	19
5.2	Logiciel de segmentation et de suivi interactif d'objets vidéo	20
5.3	Logiciel de post-production vidéo	20
6	Résultats nouveaux	21
6.1	Modélisation, segmentation et représentation d'objets Vidéo	21
6.1.1	Segmentation et suivi temporel d'objets vidéo	22
6.1.2	Schéma de codage orienté-régions optimisé pour la vidéo	23
6.1.3	Génération de maillages hiérarchiques adaptatifs et optimisation débit-distorsion pour la représentation d'images vidéo	23
6.1.4	Modélisation 3B par analyse de séquences vidéo, compression et reconstruction d'images	24
6.2	Analyse, manipulation d'objets vidéo et codage de l'information associée	25

6.2.1	Estimation du mouvement en présence de variations d'illumination . . .	27
6.2.2	Classification des différentes vues d'un objet vidéo à partir de leur mouvement	28
6.2.3	Développement de méthodes semi-automatiques pour la composition de scènes vidéo	28
6.2.4	Création d'images panoramiques	29
6.3	Couplage Source-Réseau	31
6.3.1	Couplage source-réseau pour communications sur réseaux IP	31
6.3.2	Codage conjoint source-canal pour communications sans fils	34
6.3.3	Régulation de débit pour la transmission d'images satellitaires sur canal à débit fixe	34
7	Contrats industriels (nationaux, européens et internationaux)	36
7.1	Projet européen Esprit Nemesis	36
7.2	Projet européen ACTS-Comiqs : "Commerce thru MPEG-4 on the Internet with qos"	36
7.3	Projet RNRT-Osiam	37
7.4	Projet RNRT-COHRAINTE	37
7.5	Projet RNRT-Visi	38
7.6	CTI-Cnet : Modélisation et représentation hiérarchique de scènes 3D	38
7.7	CTI-Cnet : Transmission multipoint de données multimédia sur réseaux hétérogènes	38
7.8	Contrat Celar : Liaison vidéo numérique en robotique mobile	39
7.9	Convention Cifre Microprocess	39
7.10	Convention Cifre Thomson-TCO	40
7.11	Contrat Cnes	40
7.12	Convention Cifre Thomson Multimédia	40
7.13	Convention Cifre FT-Cnet	41
8	Actions régionales, nationales et internationales	41
8.1	Actions régionales	41
8.2	Actions nationales	41
8.3	Actions internationales	42
8.4	Visites, et invitations de chercheurs	42
9	Diffusion de résultats	43
9.1	Animation de la Communauté scientifique	43
9.2	Enseignement	43
9.3	Autres actions de formation	43
10	Bibliographie	44

1 Composition de l'équipe

Responsable scientifique

Christine Guillemot [DR Inria]

Assistante de projet

Edith Blin-Guyot [TR Inria]

Personnel Inria

Claude Labit [DR, jusqu'au 1^{er} mars 1999, 1/5 temps depuis]

Henri Nicolas [CR]

Stéphane Pateux [CR, détachement du corps Télécom]

Personnel CNRS

Thierry Fauconnier [IR, (Atelier)]

Personnel Université

Luce Morin [maître de conférences, université Rennes 1]

Laurent Bonnaud [ater, université Rennes 1, jusqu'au 31 août 1999]

Collaborateurs Extérieurs

Francisco Garcia-Ugalde [PAST]

Philippe Robert [PAST]

Ingénieurs experts

Lionel Oisel [IE Inria]

François Toutain [IE Inria jusqu'au 31 mars 1999]

Chercheurs post-doctorants

Jurgen Stauder [post-doctorat européen TMR jusqu'au 31 octobre 1999]

Mustafa Sakalli [post-doctorat à partir du 1^{er} er octobre 1999]

Zhikui Chen [post-doctorat à partir du 1^{er} er décembre 1999]

Chercheurs doctorants

Dubhe Chavira-Martinez [Bourse Conacyt, à partir du 1er octobre 1999]
Franck Denoual [Bourse Cifre]
Franck Galpin [Bourse Inria-Région]
Mireya Garcia-Vasquez [Bourse Conacyt, à partir du 1^{er} septembre 1999]
Thomas Guionnet [Bourse Inria à partir du 1^{er} octobre 1999]
Xavier Henocq [Bourse Inria]
Reda Hosny [Bourse MAE]
Delphine Le Guen [Bourse Cnes]
Fabrice Leléannec [Bourse MENRT]
Gwénaëlle Marquant [Bourse MENRT]
Jean Motsch [Allocataire AMN, jusqu'au 31 août 1999]
Jean-Marie Pinel [Bourse MENRT à partir du 1^{er} octobre 1998]
Jérôme Viéron [Bourse Inria à partir du 1^{er} octobre 1998]

2 Présentation et objectifs généraux**2.1 Objectifs et Méthodologie**

Le domaine de la communication d'images a pris un envol considérable ces dernières années en liaison avec la diffusion massive d'images tant dans les domaines professionnel que domestique. Il est aujourd'hui en train de connaître des mutations importantes. Le mouvement de convergence des télécommunications, de l'informatique et de l'audiovisuel "grand public", accentué vers le milieu des années 90, accompagné de l'essor de l'Internet et des réseaux mobiles, est venu bouleverser le modèle traditionnel de la communication de données.

Ce mouvement concourt, en effet, à un changement dans les fonctionnalités à offrir dans les services de communication vidéo et multimédia. Au besoin traditionnel de communication de données ou de programmes audio-visuels, est venu se conjuguer un besoin d'intégration de média de tout type, que ce soit audio, vidéo, texte, voire même scènes de réalité virtuelle, au sein de documents dits multi - ou hyper - média interactifs, ainsi qu'un besoin d'une plus grande interactivité avec les contenus audio-visuels. Cette demande croissante de fonctionnalités nouvelles, engendre à son tour de nouveaux défis technologiques dans de nombreux domaines. Dans le domaine du traitement et de la compression du signal, il est devenu nécessaire de pouvoir représenter et coder des objets vidéos de forme arbitraire de manière à permettre une interactivité fine avec le contenu de la séquence vidéo, et de manière à faciliter l'intégration avec d'autres types de média tels que les objets de synthèse 3D.

De plus, l'expansion et la diversité des réseaux de transmission, fixes et mobiles, avec en particulier l'explosion de l'Internet, entraînent multiplication et diversification des services qui doivent faire face à une plus grande hétérogénéité des destinations, que ce soit des terminaux eux-mêmes ou des canaux de transmission reliant la source aux destinations. Ces environnements de communication hétérogènes posent de nouveaux défis de représentation et de compression du signal image et vidéo. Pour y répondre, il devient indispensable de posséder et

d'expérimenter un véritable modèle de représentation et de communication des données multimédia.

Le projet a pour objectifs de concevoir et de développer les concepts et les outils d'analyse, de modélisation, de représentation et de codage des informations vidéo manipulées en communication multimédia, permettant

- de nouvelles facilités d'interaction avec le contenu, de manipulation et d'édition de scènes, ainsi que des fonctionnalités de navigation dans des scènes vidéos 3D.
- une représentation compacte de l'information (qui reste essentielle pour de nombreuses applications telles que la vidéoconférence, les jeux vidéo ou la TV sur Internet...),
- une intégration et un mélange réaliste de plusieurs sources de données multimédia, 2D ou 3D, naturelles ou synthétiques ;
- une optimisation globale en terme de qualité de service de la chaîne de communication, prenant en compte les caractéristiques de la source, du réseau, du terminal et de l'utilisateur. Ceci devrait permettre un accès plus universel, c'est-à-dire l'accès à travers des supports de stockage ou de transmission nouveaux ou hétérogènes (Internet, Intranet, mobiles, ATM...), supposant pour certains réseaux une robustesse de l'information aux erreurs de transmission.

Nos travaux comprennent également la validation et la consolidation des concepts, des outils théoriques et des algorithmes développés, par des simulations et/ou des expérimentations au sein d'une plate-forme de communication multimédia. Le projet cherche également à valoriser les résultats de ses travaux au travers d'une part de ses partenaires industriels et d'autre part de contributions à la normalisation (IETF, MPEG).

Les nouvelles technologies numériques ainsi que l'apparition de réseaux ouverts comme l'Internet suscitent de nouveaux dangers pour la propriété intellectuelle conduisant à la réticence des possesseurs de contenus envers l'Internet, ou d'autres environnements ouverts de distribution. 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. Ceci a motivé récemment les études de marquage et d'authentification des images numériques. Les bases scientifiques du projet Temics, notamment au travers du couplage source-réseau permettent d'appréhender cette problématique. Nous pensons pouvoir apporter dans ce domaine des mécanismes conjoints marquage/codage d'images/codes correcteurs, le problème étant peu regardé sous cet angle à ce jour par la communauté scientifique et industrielle. Dans ce contexte, le projet s'intéresse plus particulièrement au marquage de séquences vidéo, en exploitant sa forte expertise en matière d'analyse, de segmentation de scènes vidéo, et de suivi des objets dans le temps.

2.2 Contexte international

Dans ce contexte, la recherche internationale s'est fédérée autour de grands thèmes tels que la communication d'images, qui nécessitent de considérer à la fois les caractéristiques des signaux de source ainsi que les caractéristiques ou contraintes des supports et services de

communication. L'ensemble de la communauté internationale académique abordant ce thème est fortement en prise avec:

- l'avancée des processus de normalisation concernant la compression et la représentation des signaux vidéo: il s'agit ici des cadres de normalisation H.26x, H32x et MPEG-1, 2, 4 ou 7; les réflexions MPEG-7 débouchent d'ailleurs sur des préoccupations plus larges que les seuls aspects de communication pour s'étendre aux outils de manipulation et d'édition de contenus vidéo, aux outils d'indexation d'images et d'aide à la navigation, à la consultation et à la manipulation de documents hypermédia
- l'avancée des processus de normalisation des protocoles de l'Internet au sein de l'IETF, et en particulier avec les évolutions envisagées vers des services intégrés (IntServ) et différenciés (DiffServ)
- le concept de médiation de la qualité de service, tant au niveau de la source que du réseau lui-même, qui tend à devenir un sujet fondamental d'étude pour une allocation optimale des ressources.
- la diversification des supports de communication (RTC, RNIS, ADSL,...), des topologies et protocoles qui leur sont associés, et avec l'hétérogénéité des qualités du service transport offert. Ce contexte est en particulier marqué par une forte expansion des réseaux sans fils pour des services mobiles qui posent de nouveaux défis en terme de représentation et de codage robuste des signaux. Ces nouveaux défis ont conduit la communauté à abandonner les principes de séparabilité du codage de source et du codage de canal inhérents au théorème de Shannon pour se tourner vers des approches de codage conjoint source-canal.

2.3 Axes de recherche

Pour répondre à ces objectifs d'identification, de représentation et de description structurée des informations vidéo manipulées en communication multimédia, dans un contexte en très forte - et permanente - évolution, nos activités de recherche sont structurées en quatre axes principaux et complémentaires :

- **modélisation, segmentation et représentation d'objets vidéo** : au sein de ce thème apparaissent naturellement les recherches en matière d'analyse du mouvement pour la compensation et l'interpolation de signaux vidéo, de description structurée, compacte et basée objets des informations vidéo, la spécification de modèles de représentations hiérarchiques (scalabilité/sélectivité) des objets vidéo ; apparaissent également les études de modélisation 3D par analyse de séquences vidéo non calibrées pour fournir une représentation compacte de scènes, appliquée à la navigation 3D ;
- **analyse, caractérisation, manipulation d'objets vidéo à des fins de composition et d'édition de scènes visuelles** : au sein de ce thème figurent les études d'analyse de mouvement, d'orientation, d'illumination, d'angles de prise de vue, de profondeur,... des objets vidéo ;
- **optimisation globale d'un service de communication multimédia par couplage source-canal ou source-réseau** : ceci induit des recherches concernant l'allocation dynamique de ressources variables, la gestion de QoS (qualité de service) adaptée aux applications traitées et la prise en compte de mécanismes de contrôle (d'erreurs ou de pertes

par congestion) pour assurer une meilleure utilisation de la bande passante disponible et ainsi une qualité vidéo restituée optimale ;

- **marquage** des contenus visuels à des fins d'**authentification robuste**, en se concentrant sur les contenus vidéo, dans la continuité de nos travaux en analyse, segmentation et suivi temporel d'une part, et de nos travaux en codage conjoint source-canal d'autre part.

Il faut également citer les concepts de **scalabilité** et de **sélectivité** ainsi que le thème sous-jacent de **représentation et de codage à réduction de débit**, qui n'apparaissent pas comme des axes de recherche à part entière mais comme des thèmes transversaux et présents dans les quatre axes principaux du projet.

3 Fondements scientifiques

3.1 Modélisation et segmentation d'objets vidéo

Mots clés : Maillages 3D, modèles par facettes, mouvement, champs denses, disparité, cartes de profondeur, géométrie projective, scalabilité, progressivité, triangulation, réalité augmentée, segmentation, analyse du mouvement, suivi temporel, modèles déformables, maillages actifs, contours actifs, codage basé régions, régions d'intérêt, théorie MDL, théorie débit-distorsion, ondelettes, échantillonnage irrégulier, quantification, maillages hiérarchiques, mosaïque vidéo, VOP, MPEG-4.

Résumé : *Dès qu'une communication d'images orientée objets vidéo est envisagée, une phase préalable d'analyse de la séquence vidéo est nécessaire. Elle consiste d'une part à segmenter les objets vidéo et d'autre part à en extraire les principaux attributs (mouvement, texture,...). Les applications visées étant la compression et la communication de données multimédia, il est également nécessaire de définir des modes de représentation compacte de la structuration des objets. Les représentations proposées doivent être faciles à manipuler (déformations géométriques ou changements photométriques) si des post-traitements de mixage vidéo (entre images réelles et synthétiques) sont envisagés. Enfin, les modèles d'analyse et de représentation spécifiés doivent permettre une description multigrille ou granulaire de la source sous formes de flux d'informations hiérarchiquement emboîtés permettant une adaptation aisée à des contraintes variables liées au canal de communication ou à l'application.*

3.1.1 Segmentation et suivi temporel d'objets Vidéo

Objectifs: L'objectif essentiel de cet axe de recherche est l'obtention d'objets vidéo possédant une localisation (spatiale ou temporelle, 2D, 2D1/2, 3D) et une caractérisation (textuelle, géométrique, topologique) les plus précises possibles. Nous nous appuyons pour ce faire sur les travaux algorithmiques de l'ancien projet Temis qui avait défini plusieurs outils complémentaires de segmentation et d'estimation de mouvement à l'aide de modèles paramétriques de

mouvement (affines bien souvent). À l'aide de ces outils d'analyse du mouvement, il convient toutefois de proposer de nouvelles stratégies permettant :

- de fournir **une meilleure localisation** (spatiale essentiellement mais également temporelle concernant les zones d'occultations) des objets segmentés. Cela passe par une prise en compte explicite (pré-segmentation spatiale) ou implicite (prise en compte de gradients locaux dans les procédures de découpage et de fusion de régions) des frontières spatiales d'objets et par un suivi temporel long-terme d'objets pour une meilleure gestion des occultations spatio-temporelles, et pour une bonne cohérence temporelle. Il est donc nécessaire de prendre en compte des caractéristiques complémentaires au mouvement telles que la forme ou la texture ;
- de fournir une **représentation compacte**, par optimisation globale du coût de description (carte de segmentation, paramètres de mouvement, texture). En effet la segmentation produite doit réaliser un compromis optimal entre une finesse de représentation et un coût limité de codage de celle-ci ;
- de proposer une hiérarchie de **cartes de segmentation topologiquement emboîtées**. Ces cartes fournissent les données d'entrée pour une approche de "scalabilité" spatiale orientée objets ;
- d'associer à ces divers niveaux de segmentation des **modes de représentation adaptés**. En effet, une représentation compacte et facilement manipulable des caractéristiques issues des phases de segmentation (frontières spatiales, contenus textuels, trajectographie temporelle,...) constitue une base indispensable pour (post-)traiter ces décompositions en objets lors des phases de restitution et de manipulation.

Approches : Nos études se concentrent tout d'abord sur la conception d'algorithmes de suivi temporel d'objets multiples - avec gestion d'occultations spatio-temporelles éventuelles - dans une séquence vidéo. Ce thème conduit naturellement à l'exploration de techniques d'interpolation temporelle utilisant les trajectoires de ces objets. L'une des difficultés essentielles réside dans la gestion des trajectoires multiples et le suivi des objets tout au long de la séquence en assurant une bonne localisation spatiale de leurs frontières ainsi qu'une bonne stabilité temporelle de la représentation et du suivi. Ceci doit s'accompagner d'une gestion locale des zones de découvrément et de recouvrement multiples qui apparaissent lors du suivi temporel des objets vidéo. Une seconde difficulté réside dans la prise en compte des déformations des objets, la plupart des techniques existantes faisant l'hypothèse d'objets rigides. Une autre difficulté consiste à représenter de manière compacte la description de la séquence analysée.

Les points essentiels de l'étude portent donc sur :

- la définition d'une représentation des cartes de segmentation spatio-temporelle adaptée à la prise en compte d'objets multiples et de leurs interactions générant des zones d'occultation et des zones de découvrément. Les techniques de segmentation pourront reposer sur des critères hybrides (texture, mouvement), sur des techniques de "mosaicking" pour gérer les zones découvertes ;
- la conception d'algorithmes de suivi temporel permettant codage et interpolation basés régions des séquences vidéo (et non pas uniquement pixels ou blocs de pixels comme dans

- le cadre MPEG actuel) ;
- la définition de modes de représentation compactes et faciles à manipuler, par optimisation globale des coûts de description, et parfois hiérarchiques. Pour ce faire, nous étudions la spécification et le suivi de maillages hiérarchiques. Les diverses motivations sous-jacentes en sont : la densité du maillage à contrôler spatialement et dynamiquement, la qualité des maillages en termes de régularité, de topologie, de stabilité temporelle, de compacité de représentation et de complexité de manipulation. Ces diverses propriétés apparaissent primordiales pour la génération de modèles scalables d'objets vidéo maillés ;
 - la conception d'un schéma de codage basé texture, mettant à profit ces diverses études et répondant aux besoins de scalabilité et de robustesse intrinsèque. Ce schéma de codage repose sur la définition d'une base de connaissances minimale (texture des différents objets présents dans une scène), permettant de reconstruire indépendamment chaque image (synthèse avec placage de textures par régions ou bien encore à l'aide de modèles déformables, puis raffinement par codage progressif de l'erreur résiduelle).

Fondements scientifiques Ces développements algorithmiques reposent sur - et contribuent à développer - les outils théoriques et bases scientifiques suivantes :

- analyse du mouvement, estimation robuste, segmentation/estimation conjointe, modélisation statistique avec régularisation, optimisation par techniques de gradient et relaxation déterministe
- segmentation hybride (texturelle/mouvement), opérateurs morphologiques
- cartes de segmentation topologiquement emboîtées
- suivi temporel d'objets, trajectoires d'objets multiples, filtrage de Kalman, interpolation temporelle basée mouvement, gestion des zones de découvrément ou recouvrement temporel d'objets
- approximations géométriques de courbes, représentation MDL ("Minimum Description Length") basée sur la théorie de l'information, codage statistique et apprentissage
- approche descendante par découpage hiérarchique de régions, approche ascendante avec relaxation et pile d'instabilité par fusions (tests de vraisemblance, approche MDL, approche débit-distorsion, métrique sur les attributs de mouvement, critère de compensation de mouvement ou de compacité de la représentation)
- représentation par maillages actifs et maillages hiérarchiques
- échantillonnage irrégulier.

3.1.2 Modélisation 3D par analyse de séquences vidéo, compression et reconstruction de vues

Objectifs : cet axe de recherche vise à définir les méthodologies et algorithmes de modélisation 3D par analyse de séquences vidéo monoculaires (une seule caméra mobile) permettant une reconstruction de vues intermédiaires pour une navigation fluide dans une scène au travers d'un réseau de télécommunications. Il s'agit également de développer des modes de représentation robuste, compacte et hiérarchique de la scène modélisée, appliqués à une navigation 3D aussi fluide que possible. Il s'attache également à exploiter les relations induites par le modèle de

projection pour exprimer les redondances entre images. L'étude se place plus particulièrement dans une perspective de compression, communication et de reconstruction de scènes, et dans un contexte de 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.

L'estimation des paramètres de modélisation 3D nécessite la mise en correspondance des images et l'estimation des matrices de passage d'un point de vue à l'autre. De nombreux travaux ont déjà été menés sur le sujet. Cependant, dans la modélisation de grandes scènes à partir de séquences vidéo monoculaires subsistent de nombreux problèmes non résolus, tels que l'estimation des paramètres "3D" de caméra, en particulier dans le cas de nombreux points de vue distants, comme ceux fournis par une caméra mobile, la cohérence d'observations entre les points de vue, la gestion des parties cachées, les régions de luminosité uniforme et les discontinuités de profondeur, etc.

En outre, au sein d'un système de navigation dans une scène au travers d'un réseau, la représentation de la scène est un point essentiel qui conditionne les performances de navigation. Elle doit permettre à la fois de restituer fidèlement la scène et d'assurer la plus grande fluidité de navigation possible. Elle contiendra au minimum des informations de texture et de géométrie, et devra permettre de conserver le réalisme de la scène, en particulier l'effet du relief durant la navigation, ainsi que le caractère spéculaire des surfaces. Il paraît donc important d'identifier de nouvelles règles de représentation et de compression, qui permettent d'optimiser le compromis entre la qualité de reconstruction et la fluidité de navigation, au travers d'un réseau. Ainsi, les différentes étapes du traitement des données sont étudiées conjointement au sein de la chaîne de reconstruction et de navigation au travers d'un réseau, afin de les optimiser globalement.

Approches : Les approches de modélisation, basées sur la mise en correspondance de points de vue distants dans les séquences, et la mise en cohérence des informations provenant des différents points de vue, reposent sur des techniques d'estimation de champs denses de mouvement respectant la contrainte épipolaire. Cette dernière est identifiée par l'extraction de points singuliers, suivie d'une mise en correspondance et d'une estimation robuste de la matrice fondamentale codant la géométrie projective. L'estimation du champ de mouvement s'appuie sur les approches robustes différentielles développées par le projet Vista pour l'obtention d'un champ régularisé avec respect des discontinuités. L'algorithme fournit ainsi un indice réel de discontinuité en chaque point du champ. L'incompatibilité de l'approche différentielle avec le traitement de points de vues éloignés (nécessaire pour une bonne estimation de la matrice fondamentale) est résolue par une approche multi-résolution.

À partir de ce champ dense de mouvement, et de cette analyse complète de la séquence vidéo, une représentation robuste de la scène, adaptée à une application de navigation 3D, est recherchée. Il faut élaborer des modes de représentation de la scène modélisée. Deux modes principaux de représentation sont étudiés : le premier consiste à approximer la scène observée par un ensemble de régions planes à partir d'un maillage initial arbitraire, de telle sorte que par exemple les facettes correspondent à un même modèle homographique de mouvement. La synthèse de nouveaux points de vues passe aussi par la connaissance des paramètres de prise de vue dans un repère euclidien, qui nécessite donc d'être capable de fournir une reconstruction

euclidienne explicite de la scène à partir du maillage 2D, sous la forme d'un maillage 3D triangulaire texturé pouvant être visualisé interactivement. Le projet se concentre aujourd'hui sur les problèmes de stabilité et de cohérence temporelle de ces reconstructions par prise en compte de la contrainte de rigidité des objets. Un second mode de représentation est une représentation basée image pour laquelle la scène est décrite par une ou plusieurs mosaïques (planaires, cylindriques, sphériques,...) auxquelles peuvent être associées des informations de profondeur. Des modes de représentations scalables selon ces deux approches sont aussi étudiés.

Fondements scientifiques : les développements méthodologiques et algorithmiques de cet axe de recherche reposent sur - et contribuent à développer - les fondements scientifiques suivants :

- estimation de champs denses de mouvement, estimation de cartes de profondeur, auto-calibration;
- reconstruction 3D par géométrie projective ;
- obtention de cartes de profondeur relative à partir de séquences non calibrées ;
- représentation par maillages actifs, représentation basée image avec des informations de profondeur associées, représentation scalable.

3.2 Analyse, manipulation d'objets vidéo et codage de l'information associée

Mots clés : mouvement, orientation, modèles d'illumination, ombrage, composition d'illumination, composition de scènes, réalité augmentée, mixage d'images naturelles et synthétiques, trajectoires 2D et 3D, images panoramiques, mosaïque d'images, édition vidéo.

Résumé : *à l'issue d'une phase de segmentation spatio-temporelle d'une séquence vidéo, nous disposons d'une vue de l'objet considéré pour chaque image originale. Cet axe de recherche a pour objectif de développer des méthodologies d'analyse, de modélisation et de représentation des caractéristiques intrinsèques de ces objets en terme de mouvement, d'angle de prise de vue, d'illumination,... afin de pouvoir recomposer ou éditer des scènes visuelles à partir des objets segmentés, et de fournir une description compacte des scènes recomposées.*

3.2.1 Objectifs

Pour permettre la composition de scènes visuelles, vidéo naturelle ou hybride, à partir d'objets 2D extraits de séquences d'images, et avec le plus grand réalisme possible, il est nécessaire d'appréhender un certain nombre de problèmes encore mal résolus :

- tout d'abord, il est nécessaire de connaître leur **mouvement 2D et/ou 3D** de manière à pouvoir les animer efficacement dans la scène mixée.
- si l'objet considéré est en mouvement relatif par rapport à la caméra (objet et/ou caméra en mouvement), alors l'objet est susceptible d'avoir été filmé sous des **angles de vue** différents. L'identification de ces angles de vues, et la classification des instances

- appelées VOP pour "video object plane" - des objets vidéo dans la séquence facilitera la composition et l'édition de scènes à partir de ces objets.
- en outre, les conditions d'éclairage des objets provenant de scènes différentes peuvent être très différentes. Si le mouvement d'un objet détermine son comportement géométrique, l'éclairage indique sa nature (apparence) photométrique. Pour la décrire, une **identification de modèles d'éclairage** (direction, intensité et couleur des sources de lumière) dans les scènes d'origine de ces objets est requise. La difficulté de ce problème réside à la fois dans la mise au point d'un modèle d'éclairage suffisamment simplifié, mais malgré tout exploitable, par rapport à la réalité physique et pouvant être efficacement utilisé dans le cas de l'analyse de séquences vidéo;
- pour le mixage vidéo, la question posée est le positionnement des objets réels ainsi que des objets synthétiques dans une scène composée. Cet objectif nécessite la prise en compte des phénomènes géométriques comme **l'occultation** entre les objets ainsi que le mouvement apparent 2D estimé lors de la phase d'analyse ;
- un autre aspect du mixage vidéo concerne les effets photométriques. Un réalisme plus grand peut être obtenu par la **génération des ombres et par la compensation des différences d'éclairage** des objets insérés dans la scène. Cela pose non seulement le problème de l'identification de modèles d'éclairage, mais également de modélisation (même très approximative) de la surface 3D de l'objet, ainsi que de sa position dans la scène.

Cet axe de recherche a aussi pour objectif l'exploitation de la connaissance des caractéristiques intrinsèques des objets vidéo pour une compression efficace des séquences originales ou recomposées.

3.2.2 Approches

Une image panoramique est définie comme l'ensemble du fond fixe de la scène vu tout au long de la séquence (dans le cas d'une caméra en mouvement). La création d'une image panoramique - technique dite de "mosaicking" - est basée sur l'utilisation de modèles de mouvement de complexité variable (modèles à trois paramètres - translation, facteur d'échelle -, ou modèle affine à six paramètres), sur des techniques d'estimation utilisant le suivi temporel. Il s'agit de définir des algorithmes permettant de calculer efficacement de telles images à partir de la séquence vidéo originale sans aucune connaissance a priori sur la scène, le mouvement et la calibration de la caméra. Les techniques de "mosaicking" permettent de gérer les découvements et les recouvrements spatio-temporels de certaines zones d'images, et d'estimer la position d'une image dans le référentiel du panoramique en évitant le passage classique par l'estimation du mouvement de la caméra.

L'éclairage dans une scène naturelle est décrite par une superposition de deux types de sources lumineuses. La première source est une source ponctuelle qui doit correspondre à l'illumination principale d'une scène. Cette source est supposée éloignée. Dans ce cas, sa description nécessite seulement un vecteur de direction avec deux degrés de liberté. Une description complète de sa position 3D correspondrait à trois degrés de liberté. Cette simplification permet de rendre robuste l'estimation des deux paramètres restants, tout en réduisant peu le réalisme du modèle. L'autre source est la lumière diffuse qui doit correspondre à toute l'illumination

sauf la source principale. La lumière diffuse est décrite par un troisième paramètre, son intensité relativement à l'intensité de la source principale. L'estimation des paramètres du modèle d'éclairage s'appuie sur une observation du signal de luminance le long de la trajectoire d'un objet en mouvement. Si l'éclairage est diffus, il y a peu de changement dans le signal ou dans l'apparence de l'objet. En présence d'une source lumineuse principale, le changement temporel du signal est utilisé pour déterminer les paramètres d'éclairage. Pour créer ce lien entre signal et éclairage, des connaissances sur le mouvement 3D et la forme 3D de l'objet en mouvement sont nécessaires. Ces connaissances sont fournies soit par l'analyse de mouvement de l'objet soit par des hypothèses statistiques comme la distribution uniforme des normales de la surface de l'objet.

3.2.3 Fondements scientifiques

Les développements algorithmiques reposent sur - et contribuent à développer - les outils théoriques et bases scientifiques suivants :

- analyse du mouvement, estimation robuste, segmentation et estimation conjointe, modélisation statistique avec régularisation, optimisation par techniques de gradient et relaxation déterministe,
- suivi temporel d'objets, trajectoires d'objets multiples, filtrage de Kalman, interpolation temporelle basée mouvement et trajectoire, gestion des zones de découvrément ou recouvrement temporel d'objets,
- étude et identification de modèles d'éclairage d'une séquence naturelle ; modélisation physique de l'éclairage, modélisation photométrique du signal d'image, modélisation des moments statistiques du signal d'une image pour l'estimation de l'éclairage, détection et suivi des ombres, analyse de mouvement 3D d'un objet par une observation de son éclairage.
- critère analytique pour l'optimisation visuelle de séquences d'images mixées.

3.3 Couplage source-réseau

Mots clés : codage conjoint source-canal, codage et décodage robustes, contrôle de congestion, codage par descriptions multiples, vidéocommunications, scalabilité, ondelettes, quantification, théorie débit-distorsion, régulation, progressivité, codes correcteurs, réseaux, internet, qualité de service, mobilité.

Résumé : *Les communications sur réseaux à qualité de service non garantie et hétérogènes ont engendré de nouveaux défis dans le domaine de la représentation et du codage des signaux audio-visuels. L'approche classique consistant à traiter de manière séparée le problème de la compression de source et celui du codage de canal, largement sous-optimale dans des environnements bruités et avec pertes, est supplantée par des approches de codage conjoint source-canal. Il s'agit alors de développer de nouvelles approches de représentation du signal vidéo et de codage conjoint de manière à améliorer les performances globales des systèmes de transmission vidéo - y compris multipoint - sur des réseaux hétérogènes.*

3.3.1 Codage conjoint pour réseaux hétérogènes en se focalisant sur les réseaux

IP

Objectifs : cet axe de recherche vise à optimiser l'utilisation de la bande passante du réseau et plus généralement à offrir des solutions au problème de la gestion de la qualité de service dans des environnements de communication hétérogènes. Il s'agit de développer des techniques de représentation et de codage du signal vidéo qui puissent s'appliquer à l'Internet 'best-effort' d'aujourd'hui, mais être également compatibles avec une évolution du réseau vers une offre de différenciation de services. Pour cela il est nécessaire d'adapter les modèles de source aux caractéristiques du canal (débit, pertes, délais,...), en introduisant des mécanismes de "couplage" entre la source et le canal. Cet axe de recherche s'intéresse à la fois à la transmission point-à-point et à la transmission multipoint sur des réseaux hétérogènes (priviliégiant l'Internet).

Approches : Le projet développe plusieurs approches complémentaires : contrôle de congestion, codage scalable à granularité fine, codage et décodage vidéo robuste, codage conjoint source-canal et codage par descriptions multiples.

La transmission de flux continus de données dans l'Internet remet d'actualité le problème classique du contrôle de congestion, dans une perspective nouvelle cependant, du fait des contraintes de délai qui interdisent l'usage des techniques classiques d'acquiescement et de retransmission. Ces mécanismes doivent permettre une bonne cohabitation avec TCP, une technique d'adaptation vivace, une bonne granularité d'adaptation des débits de la source. Cette étude considère aussi bien des réseaux de type "best-effort" que les futurs réseaux qui devraient intégrer les concepts émergents de différenciation de services (DiffServ). Les débits permis et les distorsions associées peuvent alors être liés à la priorité de service négociée. Les communications multipoints (une ou plusieurs sources vers plusieurs destinations) posent des problèmes supplémentaires liés à l'hétérogénéité des conditions de transmission dans le réseau, des capacités de décodage et de traitement dans les récepteurs. Il est nécessaire de développer des algorithmes de contrôle de congestion et de débit permettant à une source de transmettre un même flux vidéo à plusieurs récepteurs, tout en ayant la possibilité d'adapter de manière dynamique et avec une granularité suffisamment fine le débit reçu par chacun.

Les approches de représentation et de codage scalable fournissent des solutions privilégiées au problème du contrôle de congestion et d'adaptation des débits des flux aux caractéristiques non stationnaires du réseau, notamment dans un cadre de transmission multipoint ou encore d'accès à des flux pré-encodés. Un contrôle de congestion réactif et précis peut cependant exiger un nombre de niveaux de scalabilité élevé, aboutissant à des performances de compression moindres. Le projet s'attache alors à développer des algorithmes de codage scalable à granularité fine d'une part et des approches de contrôle de congestion hybrides émetteur-récepteurs d'autre part.

Le contrôle de congestion, même s'il permet de réduire les pertes dans le réseau, il ne permet pas de les éviter. Il est donc nécessaire de développer des mécanismes de contrôle de pertes, qui associés à une représentation hiérarchique des signaux, permettront de minimiser l'impact visuel des pertes. Ces mécanismes recouvrent d'une part la protection a-posteriori de l'information par des codes correcteurs adaptés, et d'autre part la robustification intrinsèque

des données par le maintien d'informations redondantes. L'un des problèmes posés par l'utilisation de codes correcteurs est le partage de la bande passante entre les données utiles et les données de redondance. Il est alors nécessaire de développer des mécanismes d'allocation conjointe de débit source et de débit de redondance, qui permettront pour des caractéristiques du réseau données, d'optimiser la qualité du signal vidéo reçu.

Le maintien d'informations redondantes dosées dans le flux compressé peut être assuré par des algorithmes de codage robuste ou de contrôle adaptatif de l'encodage de la source en fonction de caractéristiques instantanées observées au niveau de la source elle-même et au niveau du réseau, en terme de pertes, délais, gigue,... Ces techniques dites de codage conjoint source-canal regroupent à la fois les les approches d'optimisation conjointe de la quantification (scalaire et/ou vectorielle sur réseaux de points), des modes de codage prenant en compte les modèles de pertes sur le réseau, ainsi que les techniques émergentes de transformations et de quantification à descriptions multiples. Cette information redondante peut alors être exploitée par le récepteur pour dissimuler les pertes ou minimiser leur impact sur le signal reconstruit.

Notre approche consiste aussi à prendre en compte l'évolution de l'offre de service réseau soit en s'assurant d'une certaine pérennité et compatibilité des solutions développées par rapport à une évolution vers une différenciation de services, discutée au sein du groupe DiffServ de l'IETF, soit en développant les adaptations nécessaires. Le projet s'efforce également de contribuer sur les questions de transport de contenus audio-visuels sur Internet au sein du groupe AVT de l'IETF.

Fondements scientifiques : les approches développées reposent sur - et contribuent dans une certaine mesure à développer - les bases et fondements théoriques suivants :

- décomposition par bancs de filtres, par transformations en ondelettes et en paquets d'ondelettes, quantification scalaire et vectorielle multiniveaux, prédiction par EZW (*Embedded Zerotree Wavelet*), scalabilité temporelle par rafraîchissement conditionnel,
- modèles prédictifs de débit (ex : TCP-friendly), modèles de canal (ex : Elliott-Gilbert),
- modèles de régulation de débit des signaux vidéo,
- codes correcteurs, codes en blocs, FEC hiérarchiques,
- services différenciés (IETF-DiffServ).

3.3.2 Codage conjoint source-canal pour communications sans fils

Objectifs : 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 nuls. Il est donc apparu nécessaire d'optimiser les deux opérations de codage (source et canal) conjointement.

Approches : 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. L'une des questions posées est alors comment allouer au mieux le débit entre la source et la redondance pour minimiser la distorsion globale, y compris dans un contexte de codes source à longueurs variables. Dans ce 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. 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 (canal binaire symétrique, canal de Rayleigh, bruit blanc additif) variant dans le temps. On peut par exemple concevoir des transformations et des quantificateurs maintenant une quantité "contrôlée" de corrélation ou d'informations redondantes. Cette idée est à l'origine des méthodes de codage par descriptions multiples. Ces méthodes supposent l'utilisation de plusieurs canaux, et l'indépendance des erreurs sur les différents canaux, de telle sorte que la probabilité que tous les canaux soient affectés simultanément est faible. Plusieurs représentations codées du signal sont créées et transmises sur les différents canaux. Au niveau du récepteur, suivant le ou les canaux affectés par le bruit, différents schémas de décodage sont utilisés pour reconstruire le signal.

Fondements scientifiques : Les approches développées reposent sur - et contribuent à développer - les bases théoriques suivantes :

- transformations et quantifications par descriptions multiples,
- codes correcteurs d'erreurs,
- théorie débit-distorsion,
- représentation hiérarchique - ou scalable - des sources, décomposition par transformations en ondelettes ou en paquets d'ondelettes, quantification algébrique emboîtée, prédiction de type EZW,
- quantification scalaire, vectorielle, algébrique,

3.4 Tatouage d'images et de contenus vidéo

Mots clés : tatouage (watermarking), authentification, marquage robuste.

Les bases scientifiques du projet Temics, au travers de son expertise en analyse spatio-temporelle des séquences vidéo et en matière de codage conjoint source-canal ont conduit naturellement le projet à s'intéresser au problème du tatouage de séquences vidéo, à des fins d'authentification des contenus. Nous nous intéressons plus particulièrement aux séquences vidéo, en développant des mécanismes conjoints de marquage, de codage d'images et de codes correcteurs. Plus précisément, nous nous intéressons à la robustification de la marque par l'utilisation de codes correcteurs, à l'enrichissement du décodage de la marque par une mesure de confiance dans les observations, à l'utilisation du codage par descriptions multiples pour un

tatouage robuste aux attaques, à l'insertion de marques à des positions temporellement stables afin d'éviter des attaques par filtrage temporel.

3.5 Thèmes transversaux: scalabilité et sélectivité

La notion de **scalabilité** - anglicisme désignant une structuration multiéchelle des données - apparaît désormais incontournable pour représenter et communiquer efficacement des images (communication multicast, réseaux à caractéristiques variables et hétérogènes). Cette notion se décline selon plusieurs aspects complémentaires :

- **la scalabilité objets** : les techniques de segmentation spatiotemporelle hiérarchique conduisent naturellement à une représentation multiéchelle des objets segmentés ;
- **la scalabilité spatiale** : la représentation multirésolution du signal ainsi que la découpe ordonnée en objets de profondeurs distinctes (segmentation multiniveaux) sont également des modes aisés de structuration du signal vidéo ;
- **la scalabilité temporelle** : en extension de la notion de GOP ("Group of Pictures") classique dans le cadre MPEG qui structure le flux vidéo en trames IBP, des modes étendus de découpe temporelle sont à définir ;
- **la scalabilité de représentation** : ceci concerne la représentation d'une source d'images en un ensemble ordonné de sources emboîtées permettant la restitution partielle ou progressive des images. D'un point de vue du flux binaire, ceci se traduit par un ensemble de flux élémentaires à concaténer.

Il convient d'adjoindre à cette décomposition multiéchelle des données vidéo, une notion de **sélectivité**. En effet, il apparaît de plus en plus souhaitable, pour certains services de communication (télémédecine, imagerie satellitaire, vidéosurveillance, robotique mobile) de modéliser et d'introduire les connaissances a-priori dans l'algorithmique de représentation et de compression. Cette sélectivité est définie à partir de modèles perceptuels ou est guidée par l'application et le contenu de scènes, ce qui conduit à définir des régions d'intérêt. La sélectivité s'exprime alors par une hiérarchisation de l'information tant spatiotemporelle (basée sur le mouvement) que fréquentielle (basée sur des lois de sensibilité en fréquences).

4 Domaines d'applications

4.1 Panorama

Nous nous intéressons à trois grands types d'application. Le premier concerne la compression d'images fixes et animées. Nous explorons dans ce cadre les domaines de la compression d'images satellitaires (avec le Cnes), et de la compression scalable du signal vidéo (images de visioconférence avec le Cnet). 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). Le dernier secteur a trait à la production et à la post-production vidéo. Les domaines d'application sont les outils de création de contenus multimédia.

4.2 Compression d'images fixes et animées

Les cadres applicatifs naturels de ce thème concernent, pour une part, les aspects de stockage (compression) et de consultation aisée (décompression) de bases de données d'images et d'autre part, la transmission de signaux (audio)visuels sur des à bande étroite. Les outils de compression d'images existants répondent désormais à la majeure partie des applications "standards" pour le codage et le décodage source pour des moyens et hauts débits avec qualité de reconstruction homogène au sein de la scène visuelle. Il demeure par contre de nombreuses investigations à mener concernant les applications de compression d'images fixes Haute-résolution, ceux de compression quasi sans pertes avec prise en compte de modèles basés psychovision et ceux de compression sélective. Dans ce domaine, les modalités d'imagerie satellitaire (que nous traitons en collaboration avec le Cnes) posent des questions particulières de qualité de reconstruction et de caractérisation des dégradations. Enfin, mentionnons les recherches toujours nécessaires pour étendre les outils de la théorie de la quantification et du codage entropique aux applications de codage bas-débits (inférieures à 0.5bpp).

4.3 Communication multimédia

Les services et les applications multimédias qui sont en cours de déploiement évoluent en permanence et demandent à être enrichis. Les nombreuses actions de développement qui ont pour finalité la définition des services de consultation, de diffusion et de communication interactive multimédia reposent, pour la représentation, la manipulation et la présentation des contenus, sur l'utilisation de normes principalement orientées "compression". Les technologies qu'elles mettent en œuvre présentent certaines limitations : elles permettent des degrés limités d'interaction avec le contenu de l'information multimédia et sont très sensibles aux pertes ou aux erreurs de transmission.

Les communications audiovisuelles interpersonnelles, la diffusion, le stockage et la consultation de documents audiovisuels dans des environnements hétérogènes fixes ou mobiles représentent ainsi un domaine d'application important de nos travaux algorithmiques. Noter que la transmission multipoint dans des environnements hétérogènes s'avère un besoin de bon nombre d'applications communicantes telles que la visioconférence, le travail coopératif, ou la diffusion.

Les domaines applicatifs plus particulièrement traités par le projet, au travers de ses partenariats industriels bi-latéraux et européens concernent tout d'abord les applications de consultation multimédia sur Internet (télé-achat, télévisite, téléenseignement). Dans ce type d'application, le codage des contenus multimédia (audiovisuels, 3D synthétique) est en général réalisé en différé, ce qui conduit à un service asymétrique où seul le décodage doit être réalisé en temps réel et 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 constitue un point de rencontre avec le domaine de l'indexation, de la consultation et de la gestion de bases de données vidéo.

Ils concernent ensuite les communications de groupes (ex: visioconférence) que nous traitons en collaboration avec le Cnet et au sein du projet RNRT Visi. Ils concernent aussi les applications de diffusion de compte-rendus médicaux et de télédiagnostic coopératif, que nous abordons au travers de notre participation au groupe de réflexion Thélème et de notre colla-

laboration avec l'Université de Brescia sur le codage avec régions d'intérêt pour le télédiagnostic médical. L'aspect consultation d'images par le praticien devient un enjeu important et critique en termes de ressources réseaux et en terme de qualité de service. Le cadre de normalisation MPEG-7, traitant simultanément des aspects de consultation (indexation) et représentation (compression) des données multimédia, constitue un point de rencontre entre ces différents domaines. Enfin ils concernent les vidéocommunications mobiles que nous traitons en collaboration avec le Celar au travers d'une application de téléguidage d'un robot terrestre mobile.

4.4 Production et post-production vidéo

La nécessité de disposer d'outils automatiques ou semi-automatiques (l'aspect interactif via des IHM adaptées n'est pas à exclure dans un tel cadre) permettant la manipulation d'objets vidéo apparait de plus en plus fortement dans le domaine de la post-production vidéo, notamment pour la création d'effets spéciaux et la composition de scène (cadre MPEG4). Effectivement, si les outils de post-production actuels permettent de mettre en œuvre de telles modifications sur les données vidéo originales, ils sont très largement manuels et, de ce fait, leur mise en œuvre est difficile et coûteuse. Nous pouvons distinguer deux types fondamentaux de manipulations d'une séquence vidéo acquise par une caméra : l'édition vidéo pour laquelle on souhaite supprimer des objets dans la scène ou ajouter de nouveaux objets réels ou synthétiques, et la modification artificielle de la position, de la focale ou du mouvement de la caméra permettant de changer a posteriori le cadrage de la scène en fonction de l'application visée. Ces problèmes sont spécifiquement traités au sein du projet Esprit Nemesis.

À cela viennent s'ajouter les applications nécessitant la mise en œuvre de techniques de réalité augmentée, comme les applications de mondes virtuels partagés, de téléprésence, jeux,... e concept de réalité augmentée se traduit essentiellement par l'hybridation possible de scènes ou objets issus du monde réel (via vidéo) et issus de modeleurs synthétiques. La difficulté majeure est alors la détection et la segmentation finement localisées des objets vidéo considérés et la représentation efficace de ceux-ci assurant, par exemple, une compatibilité avec les modeleurs géométriques et radiométriques d'objets de synthèse. Par ailleurs, la modification des caractéristiques du capteur, en simulant par exemple un capteur virtuel, nécessite également une modélisation géométrique d'ordre supérieur à la simple projection perspective bidimensionnelle sur un plan image ; ceci recherche l'identification de plans de profondeurs et vise la création de vues intermédiaires par interpolation, pour de la navigation dans une scène au travers d'un réseau, rejoignant en cela le domaine de la communication multimédia. Ce domaine constitue le cadre de nos travaux au sein du projet RNRT-V2NET.

5 Logiciels

5.1 Plate-forme de communication multimédia

Participants : Thierry Fauconnier, Christine Guillemot, Reda Hosny, François Toutain, Jérôme Viéron.

Mots clés : compression, communication sur réseaux IP, fonctionnalités objets, interactivité basée contenu, multicast.

Pour faire évoluer le domaine et ses applications, il devient indispensable de globaliser des objectifs multidisciplinaires, et en particulier d'aborder avec une plus grande cohérence les différents éléments d'une chaîne de traitement et de transmission de l'information. Ceci nécessite de considérer et d'appréhender dans sa globalité le problème de la représentation et du codage des signaux et celui du transport de ces signaux. La plate-forme de communication multimédia, en cours de développement, s'inscrit dans cette démarche. Elle a pour but de fédérer, en proposant un cadre d'expérimentation et de validation des algorithmes, les axes de recherche du projet. Elle devrait, à terme, permettre le test en "conditions réelles" les algorithmes développés. L'objectif est en effet de permettre d'émuler une vaste gamme de conditions de réseaux (congestions, délais, pertes). Partant d'une configuration point à point en mode consultation, réalisée en 1999 au travers de notre participation au projet Européen ACTS-Comiqs, la plate-forme sera étendue vers une configuration multicast puis vers une configuration distribuée hétérogène, dans les projets RNRT-VISI et IST-SONG, en l'an 2000. Elle sera raccordée à la plate-forme IPV6 et de test de différenciation de services du projet Armor.

5.2 Logiciel de segmentation et de suivi interactif d'objets vidéo

Participant : Stéphane Pateux.

Mots clés : segmentation, suivi, masque d'objet.

Dans le cadre du projet Osiam, a été développée une interface graphique (voir fig.1) permettant la segmentation interactive d'objets vidéo. La partie interface de ce logiciel a été écrite en Java, ce qui permet une utilisation multi-plate-forme de l'outil. Le développement d'un tel outil répond au besoin de segmentation en objets qui apparaît de plus en plus, notamment pour les applications de compression (codage orienté objets MPEG4), ou bien encore de post-production. Dans un premier temps, l'utilisateur réalise une segmentation de la première image par regroupement à l'aide de la souris de régions obtenues par une segmentation spatiale. Le logiciel permet ensuite de réaliser un suivi des masques de segmentation définis image par image. A tout moment, l'utilisateur peut intervenir afin de corriger les éventuels défauts rencontrés lors du suivi. A terme, il est prévu de faire évoluer cet outil en y intégrant d'autres outils de segmentation et de suivi (segmentation spatio-temporelle, suivi par contours actifs) afin de répondre au mieux à la diversité des situations rencontrées.

5.3 Logiciel de post-production vidéo

Participants : Henri Nicolas, Lionel Oisel.

Mots clés : post-production, images panoramiques, segmentation statique, segmentation dynamique, recalage 3D/2D.

L'objectif du projet européen Nemesis (voir section 7) est de développer et de commercialiser un logiciel de post-production vidéo s'adressant tant aux professionnels du domaine qu'aux amateurs avertis. Le choix de la plate-forme de développement s'est orienté vers l'intégration des outils développés dans le cadre de Nemesis à un logiciel de post-production déjà largement diffusé, «Adobe After Effect». Outre sa large diffusion, ce dernier présente l'avantage d'offrir

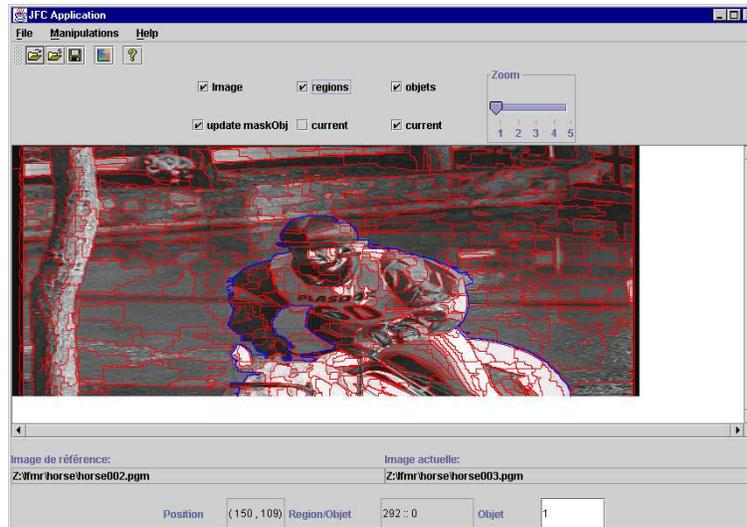


FIG. 1 – Interface graphique pour la segmentation interactive d’objets vidéos. A l’aide de la souris, l’utilisateur peut regrouper les régions d’une segmentation spatiale (frontières en rouge) afin de définir et/ou de corriger les objets vidéos extraits (frontières en bleu).

une interface dotée de nombreux outils basiques de gestion de séquences vidéo. Les algorithmes développés dans Nemesis se présentent donc sous la forme de fonctions optionnelles (appelées «plug-ins») qui utilisent l’interface d’«After Effect». Le projet Nemesis a alors développé des outils de segmentation statique (laboratoire ICS, Grèce), de segmentation dynamique (laboratoire HHI, Allemagne), de recalage 2D/3D et étalonnage interactif (Inria Rocquencourt, France), et de création d’images panoramiques (Irisa-projet Temics). Une version beta destinée à être testée par des professionnels de la post-production est en cours de réalisation.

6 Résultats nouveaux

et représentation d’objets Vidéo

6.1 Modélisation, segmentation et représentation d’objets Vidéo

Mots clés : Maillage 3D, modèles par facettes, mouvement, champs denses, disparité, cartes de profondeur, géométrie projective, scalabilité, progressivité, triangulation, réalité augmentée, segmentation, analyse du mouvement, suivi temporel, modèles déformables, maillages actifs, contours actifs, codage basé régions, régions d’intérêt, théorie MDL, théorie débit-distorsion, ondelettes, échantillonnage irrégulier, quantification, maillages hiérarchiques, mosaïque vidéo, VOP, MPEG-4.

6.1.1 Segmentation et suivi temporel d'objets vidéo

Participants : Laurent Bonnaud, Stéphane Pateux.

Dans le cadre du projet RNRT Osiam, une nouvelle technique de suivi d'objets vidéos a été développée, permettant de répondre au besoin d'applications diverses (codage MPEG4, surveillance,...), les objets suivis pouvant être quelconques et notamment déformables. Cette technique est basée sur la projection en arrière d'une segmentation spatiale (projection backward). Le principe du suivi est d'effectuer pour chaque image une segmentation spatiale de celle-ci, puis, par l'intermédiaire d'une projection vers le passé, au sens du mouvement, sur une carte de segmentation en objets de référence, de trouver à quel objet appartient cette région. L'intérêt de ce procédé est de fournir ainsi une bonne localisation des frontières des objets suivis, et que de traiter les objets déformables qui sont généralement difficile à suivre. Par ailleurs, la complexité de cette méthode de suivi permet d'envisager des implémentations temps réel de cette technique de suivi. Toutefois, il peut arriver qu'une région spatiale se projette sur plusieurs objets définis par le masque de référence (cas des zones d'occultation, d'erreurs dans la segmentation spatiale). Une technique de redécoupage au sens du mouvement des régions spatiales a alors été développée afin de palier ces problèmes. Afin de simplifier l'utilisation de cet outil de segmentation et de pouvoir interagir sur les résultats obtenus, une interface graphique a été réalisée. Cette interface permet ainsi de définir le masque initial des objets, de lancer le suivi, et de corriger les dérives éventuelles de l'algorithme de suivi.

Parallèlement à cette étude de suivi d'objets vidéo, un nouvel outil de segmentation d'images fixes multi-composantes a été développé. Cet outil de segmentation est basé sur le formalisme MDL, et permet d'obtenir une de segmentation précise avec peu d'interaction. Cette nouvelle technique de segmentation s'avère être extrêmement intéressante dans le cas de segmentation d'images couleurs. En effet, la segmentation est indépendante de l'espace linéaire de représentation des couleurs (RGB, YUV,...), et contrairement aux techniques de segmentation par clustering permet de ne pas sur-segmenter les régions ayant un dégradé de couleur. Ces résultats sont illustrés à la fig.2.



FIG. 2 – Suivi temporel d'objets vidéo. (a). Segmentation itiale (image 61); (b). suivi pour l'image 65; suivi pour l'image 90 (c). Malgré une forte déformation de l'objet et un mouvement de la caméra, l'objet considéré est correctement suivi, et les frontières de la région restent précises.

6.1.2 Schéma de codage orienté-régions optimisé pour la vidéo

Participants : Stéphane Pateux, Claude Labit.

Mots clés : codage, optimisation débit-distortion, segmentation.

Dans un schéma de codage orienté-régions de séquences vidéo, le choix de la segmentation à utiliser pour le codage est un sujet d'études essentiel. Les travaux précédemment menés au sein du projet étaient issus de travaux d'analyse du mouvement $2D+t$ qui n'avaient pas été forcément adaptés en terme d'efficacité de codage. Dans la thèse de S. Pateux, nous avons utilisé le formalisme MDL (Minimum Description Length), afin d'obtenir une segmentation adaptée au codage d'images. Cependant cette technique ne permettait pas aisément de réaliser un codage à qualité constante ou bien encore à débit fixé. Une approche d'optimisation selon un critère débit-distorsion a été conçu afin de pouvoir répondre à cet objectif.

La recherche de la segmentation optimale est basée sur une technique d'étiquetage d'une sur-segmentation initiale d'une image. L'étiquetage est réalisé par l'intermédiaire d'opérations de fusion de régions et de relaxation d'étiquettes attachées aux régions. L'ordonnancement de ces opérations est contrôlé à l'aide d'un critère de type débit-distorsion. Ce même critère est utilisé pour sélectionner le meilleur mouvement pour une région issue d'une fusion de deux régions, parmi plusieurs mouvements proposés. Additionnellement, une variation du pas de quantification est mise en concurrence avec les opérations de fusion et de relaxation sur les étiquettes des régions. Les opérations de fusion, de relaxation, et la variation du pas de quantification sont menées jusqu'à l'obtention du résultat souhaité (débit cible atteint ou bien qualité désirée obtenue).

Les tests réalisés sur plusieurs séquences vidéo montrent un gain significatif par rapport aux normes de codage existantes (MPEG1,2 pour la vidéo et H261,3 pour la visiophonie). Ce gain se traduit en terme de rapport signal à bruit (gain de l'ordre de 1 à 2 dB par rapport aux normes de codage les plus récentes), mais également en terme de qualité visuelle (limitation des effets de bloc, stabilité temporelle de la qualité, netteté des images même en présence de mouvement).

6.1.3 Génération de maillages hiérarchiques adaptatifs et optimisation débit-distorsion pour la représentation d'images vidéo

Participants : Gwenaëlle Marquant, Stéphane Pateux.

Mots clés : Échantillonnages régulier et irrégulier, maillages actifs, mouvement, maillage hiérarchique, scalabilité.

L'objectif de cette étude est de générer des maillages triangulaires scalables pour le codage d'images fixes. A partir des travaux de thèse de Hervé Le Floch qui s'appuyaient sur un échantillonnage irrégulier pour représenter les images fixes et animées, nous sommes arrivés à la conclusion d'une nécessaire adaptation de la densité des noeuds au contenu de l'image. Cependant, d'autres contraintes (comme le coût de codage par exemple) nous incitaient à utiliser des

structures régulières. Il a donc fallu trouver un compromis : un maillage triangulaire régulier hiérarchique adapté aux caractéristiques de l'image. La hiérarchie, basée sur une technique itérative générale de division, est construite à partir de l'utilisation d'un critère débit-distorsion. Le choix de ce critère adéquat, simple, et peu coûteux prend en compte les aspects de débit et de distorsion, ce qui nous permet de contrôler les coûts de codage.

Parallèlement au développement de critères permettant d'optimiser la densité du maillage, des techniques spécifiques d'évaluation et de codage des valeurs aux noeuds ont été développées. Ainsi, dans le cas du codage d'images fixes, une optimisation des valeurs de luminance (et de couleur) aux noeuds a été développée, et une étude a été menée sur les techniques de quantification adaptées au codage d'images par maillage hiérarchique. Le schéma ainsi proposé permet de réaliser un codage progressif d'une image efficace avec une granularité fine.

Pour des séquences d'images, une technique d'estimation de mouvement sur des maillages hiérarchiques topologiquement emboîtés a été développée, offrant plusieurs niveaux de représentation des déformations d'un objet. Les travaux actuels portent sur l'optimisation de la densité locale de noeuds du maillage pour la compression de séquences vidéo, le suivi d'objet vidéo et sur la prise en compte des discontinuités présentes dans l'information de mouvement (zone d'occultations, objets multiples).

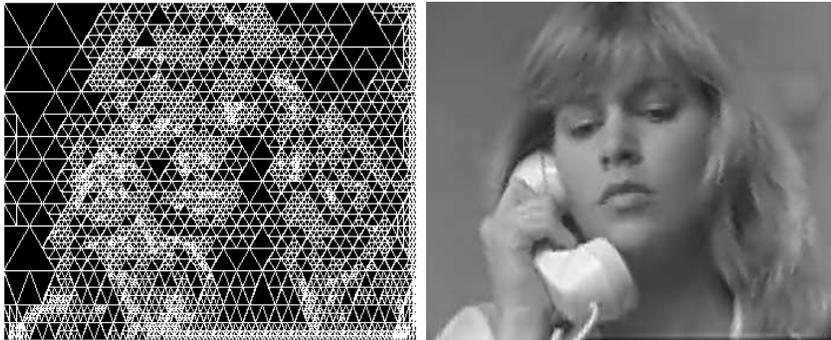


FIG. 3 – *Codage d'images fixes par maillages. À gauche grille adaptée obtenue à partir de 1600 divisions, soit 0.22 bits/pixel; À droite image reconstruite par interpolation affine et optimisation des valeurs de niveaux de gris aux noeuds, 1600 divisions, 31.65 dB après optimisation (30 dB sans optimisation).*

6.1.4 Modélisation 3B par analyse de séquences vidéo, compression et reconstruction d'images

Participants : Franck Galpin, Luce Morin, Philippe Robert.

Mots clés : Maillages 3D, modèles par facettes, mouvement, champs denses, disparité, cartes de profondeur, géométrie projective, scalabilité, progressivité, triangulation, réalité augmentée.

Les études réalisées cette année ont porté sur la réalisation d'une chaîne complète permettant une reconstruction projective ou pseudo-euclidienne d'une scène rigide observée depuis deux points de vue. La méthode proposée comprend les étapes suivantes :

- une estimation robuste de la matrice fondamentale à partir de paires de points caractéristiques préalablement extraits et mis en correspondance,
- une estimation du champ dense de disparité contrainte par la géométrie épipolaire, obtenue par un algorithme d'estimation robuste et multi-résolution de flot optique,
- une segmentation en facettes planes de la scène à partir de l'information 2D du flux optique,
- une reconstruction 3D de la scène sous la forme d'un modèle VRML texturé.

La contrainte épipolaire a ensuite été introduite dans un algorithme de segmentation et d'estimation conjointes de mouvement : l'évaluation comparative des résultats est en cours. La reconstruction de la scène a également été améliorée en introduisant une segmentation spatiale dans l'estimation du champ de disparité, une comparaison d'approches 2D et 3D pour la segmentation en facettes planes, ainsi qu'une méthode permettant de compenser les distorsions perspectives induites dans l'opération d'habillage de l'objet par plaquage de textures prises dans une ou plusieurs images (voir fig.5). Il est en effet nécessaire de roduire une texture vue de face pour chaque facette triangulaire du modèle 3D. La méthode développée permet de fournir une seule image pour tout l'objet, dans laquelle les distorsions perspectives de textures sont corrigées, plutôt qu'une image par facette.

Une première généralisation de l'algorithme à une séquences d'images (en considérant plusieurs images successives) a été réalisée. Elle comprend :

- l'autocalibration des paramètres intrinsèques de la caméra,
- le suivi du mouvement par estimation et cumul d'un champ dense non contraint,
- la modélisation de la séquence sous la forme de plusieurs modèles 3D possédant des points de vue communs.

Il est ainsi possible de générer à partir de ce modèle toute séquence synthétique dont les vues sont proches de la séquence originale. La spécification de la séquence virtuelle est effectuée à l'aide d'une interface graphique qui permet de placer la caméra virtuelle par rapport au modèle 3D, et de définir ainsi la séquence virtuelle par un ensemble de positions clés (voir fig.4). Ce travail se poursuit dans le cadre du projet RNRT-V2NET.

6.2 Analyse, manipulation d'objets vidéo et codage de l'information associée

Mots clés : mouvement, orientation, modèles d'illumination, ombrage, composition d'illumination, composition de scènes, réalité augmentée, mixage d'images naturelles et synthétiques, trajectoires 2D et 3D, images panoramiques, mosaïque d'images, édition vidéo.

Résumé : *L'objectif général est ici de développer des techniques permettant la manipulation basée objet de séquences vidéo originales pour des applications liées à la post-production vidéo ou la composition de scènes. Par manipulations basées objet, nous entendons ici l'ajout ou la suppression d'un ou plusieurs objets vidéo, la*

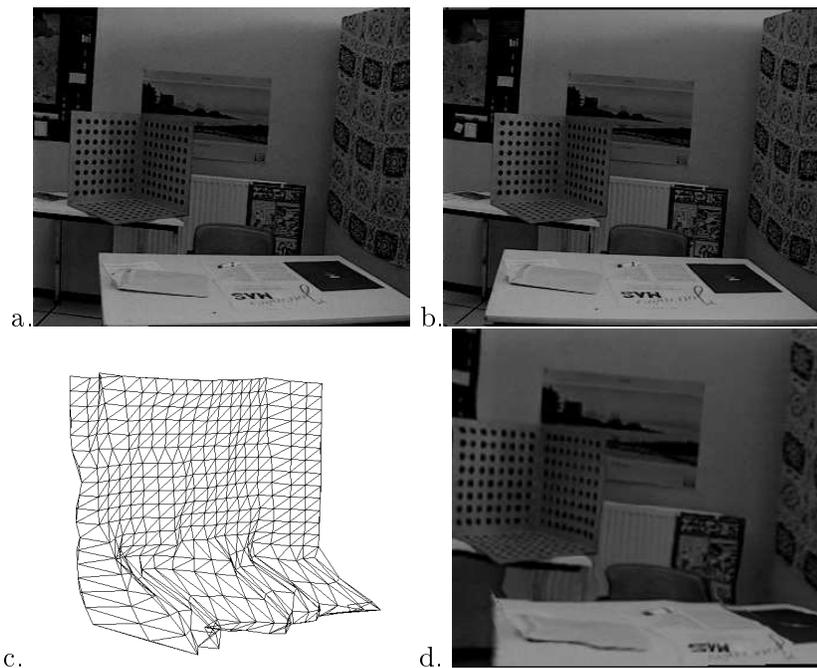


FIG. 4 – *Reconstruction pseudo-euclidienne à partir d'images: (a) et (b) images originales 1 et 50 de la séquence "armel"; (c) modèle 3D: vue latérale du maillage et vue synthétique de la scène; (d) vue synthétique de la scène, obtenue par projection du modèle texturé*

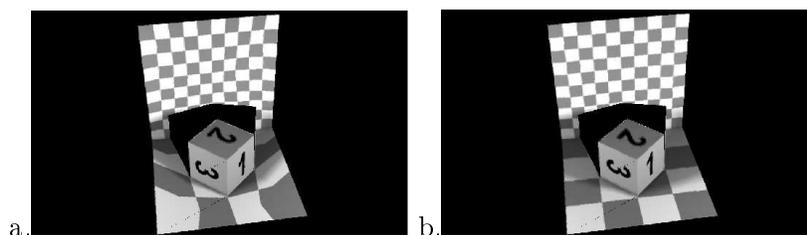


FIG. 5 – *Recalage de textures réelles pour l'habillage de modèles 3D: (a) modèle 3D reconstruit à partir de deux images de synthèse; (b) sans recalage de texture (à gauche), avec recalage de texture (à droite)*

modification de la trajectoire d'un objet, ou encore la création d'images mosaïques. La mise en oeuvre efficace de ce type de traitement, nécessite souvent l'analyse des caractéristiques de mouvement et des conditions d'éclairage des objets vidéo. De plus des méthodes sont développées de manière à automatiser au maximum le processus de composition de scène.

6.2.1 Estimation du mouvement en présence de variations d'illumination

Participants : Jurgen Stauder, Henri Nicolas, Jean-Marie Pinel.

Mots clés : modèles d'illumination, ombrage, estimation du mouvement, compensation d'illumination.

Cette étude a pour objectif la segmentation de séquences vidéo (extraction des VOP) par la caractérisation du mouvement des objets grâce à la prise en compte des variations d'ombrage générées par ce mouvement. Il a été montré qu'il est possible d'estimer la projection - dans le plan image - de l'angle de rotation 3D d'un objet illuminé par une source ponctuelle en utilisant simplement ses variations d'illumination (voir fig.6). Cette méthode a été appliquée à la segmentation d'une séquence de VOPs en fonction des variations de l'angle de rotation estimé. Pour cela, les segments temporels sont identifiés en fonction de l'évolution temporelle de l'angle de rotation estimée par un filtre de Kalman. Elle est directement complémentaire avec la méthode de classification (basée mouvement) des VOPs décrite au paragraphe suivant. Cette méthode a été testée sur des séquences de type vidéoconférence pour lesquelles des résultats satisfaisants ont été obtenus.



FIG. 6 – Deux segments temporels de mouvement homogène trouvés automatiquement pour la séquence “Tai”, format Cif, fréquence 12.5Hz: (a) images 16 ... 31 et (b) images 36 ... 50. Dans les images, la projection 2D de l'axe de rotation 3D de l'objet est représentée.

6.2.2 Classification des différentes vues d'un objet vidéo à partir de leur mouvement

Participants : Jean Motsch, Henri Nicolas, Mireya Garcia-Vazquez.

Mots clés : classification, mouvement, VOP, mixage, réalité augmentée, mixage d'images réelles et synthétiques, trajectoires 2D et 3D, suivi.

Nous avons cherché à obtenir une classification des VOP selon un critère basé sur leurs angles de prise de vue respectifs (l'objet est supposé rigide et sans variations d'illumination). En pratique, si l'objet est animé d'un mouvement translationnel 3D ou de rotation dans le plan image, l'angle de vue de l'objet peut être considéré comme constant (si l'amplitude du mouvement de translation n'est pas trop importante en fonction de la distance objet-caméra, et des variations relatives de profondeur de l'objet considéré). A l'inverse, en cas de rotation autour d'un axe parallèle au plan image, l'angle de vue se modifie. La méthode développée utilise un critère de vraisemblance permettant d'évaluer le caractère significatif ou non des paramètres d'un modèle de mouvement affine utilisé pour estimer et décrire le mouvement relatif entre l'objet et la caméra. Sur cette base, chaque transition entre deux VOP successifs est étiquetée comme constante (pas de modification de l'angle de vue) ou variable (modification du point de vue). Une sous-classification plus fine est également introduite en associant à chaque transition le type de mouvement correspondant. L'objet vidéo est alors vu comme une succession de segments temporels avec ou sans modification du point de vue. De plus, tous les VOP d'un même segment constant peuvent être considérés comme similaires (si l'on ne prend pas en compte les effets d'illumination ou d'occultation). Néanmoins, chacun d'entre eux peut avoir un facteur de zoom différent, et donc une résolution spatiale, différent. Le VOP ayant la résolution spatiale la plus forte est donc celui qui représente le mieux le segment constant considéré.

6.2.3 Développement de méthodes semi-automatiques pour la composition de scènes vidéo

Participants : Franck Denoual, Henri Nicolas.

Mots clés : composition de scènes, mixage.

Nous développons de nouvelles méthodes semi-automatiques de création de séquences d'images mixées à partir de plusieurs objets vidéo ou de mosaïques. Pour cela, il est nécessaire de définir, dans la scène mixée, la position de chacun des objets pour chacune des images créées sur la base d'un scénario de composition préalablement défini par l'application. Dans le domaine de la post-production vidéo, ce type de mixage est réalisé manuellement, même si des interfaces conviviales existent pour faciliter le travail des opérateurs. Notre objectif est de réduire au minimum l'intervention humaine en automatisant au maximum la procédure de mixage. La trajectoire de chacun objet inclut dans la séquence mixée est découpée en un ensemble de segments dans lesquels la trajectoire de l'objet peut être considérée comme linéaire. Ces instants clés sont obtenus automatiquement par l'analyse des trajectoires originales et mixées de l'objet, bien que la possibilité soit offerte à l'utilisateur de rajouter manuellement

d'autres instants clés. L'opérateur doit alors positionner manuellement chaque objet vidéo pour chacun de ces instants clés, la position des vues intermédiaires étant calculée automatiquement par interpolation. De plus, une méthode basée sur la notion de lignes de niveau permet de fixer automatiquement le facteur d'échelle de l'objet en fonction de sa profondeur théorique dans la scène (une ligne de niveau correspond, pour une image donnée, à l'ensemble des points se situant à une distance identique de la caméra). Si l'on dispose d'une image mosaïque représentant le fond de la scène mixée, ces lignes sont calculées de manière semi-automatique en début de traitement. Sinon, les lignes sont calculées image par image à partir du mouvement de la caméra. Cette méthode a été validée sur des séquences de test réelles.

6.2.4 Création d'images panoramiques

Participants : Henri Nicolas, Lionel Oisel.

Mots clés : images panoramiques, mosaïque d'images, édition vidéo, manipulation d'images, post-production vidéo.

L'objectif de cette étude est de définir une méthode de calcul efficace d'images mosaïques à partir de la séquence vidéo originale sans aucune connaissance a priori sur la scène, le mouvement et la calibration de la caméra. Nous estimons directement la position d'une image dans le référentiel du panoramique, en évitant le passage classique par l'estimation du mouvement de la caméra, tout en cherchant à minimiser le temps de calcul. En effet, dans le cadre des applications de post-production vidéo, bien qu'il ne soit pas nécessaire de travailler en temps réel, il est indispensable de garder une complexité opératoire raisonnable. La technique d'estimation multirésolution du mouvement développée ici permet de calculer une image mosaïque (format Cif) en 3-4 secondes/image originale sur PC. Cette technique utilise des modèles de mouvement de complexité variable, un suivi temporel et la mise en correspondance de chacune des nouvelles images originales avec l'image mosaïque courante. De très bons résultats ont été obtenus sur des séquences complexes ayant un mouvement de caméra rapide (par exemple la séquence "stefan") et sur des séquences de test issues de la post-production vidéo (voir fig.7).

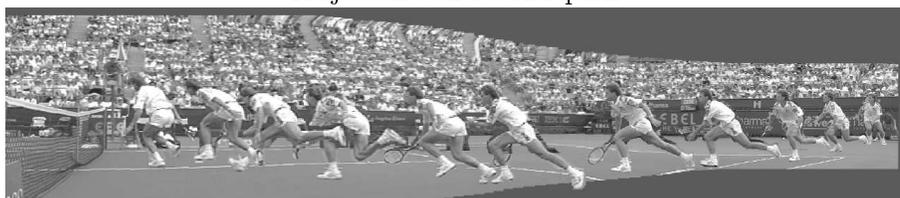
Un deuxième aspect important concernant les techniques de création d'images panoramiques est lié à la manière de gérer le recouvrement des images. La technique classique consiste soit à remplacer l'ancienne information existante dans l'image panoramique par la nouvelle image, soit au contraire à ne compléter l'image panoramique que là où elle n'est pas encore définie. Cette technique n'est pas optimale car elle ne tient pas compte des différentes résolutions de l'image (provenant des variations de focale de la caméra), de l'ancienneté de l'image panoramique, ainsi que des effets de faux contours aux niveaux des raccordements entre les images. Pour prendre en compte ces différentes contraintes, un critère original permettant d'optimiser efficacement le choix de coefficients de pondération entre la nouvelle image et l'image panoramique existante a été développé et intégré au schéma d'estimation. De plus, le calcul des images panoramiques dépend de deux paramètres d'optimisation correspondant respectivement à la contrainte de résolution et au délai temporel. L'opérateur peut donc choisir les valeurs de ces paramètres en fonction des contraintes de l'application. Cette technique a été optimisée et a été implantée au sein du logiciel Nemesis dont la version *beta* a été finalisée en automne 99.



Scène originale



Trajectoire en second plan



Trajectoire au premier plan



FIG. 7 – Exemples de modification de la trajectoire d'un objet vidéo. De haut en bas: trajectoire originale, trajectoire de l'objet mise au second plan, à l'avant-plan, et trajectoire inversée. Les rectangles indiquent les VOP-clés ajoutés à la main

6.3 Couplage Source-Réseau

Mots clés : codage conjoint source-canal, codage/décodage robuste, contrôle de congestion, codage par descriptions multiples, vidéocommunications, scalabilité, ondelettes, quantification, théorie débit-distorsion, régulation, progressivité, codes correcteurs, réseaux, internet, qualité de service, mobilité.

Résumé : *Les travaux ont porté sur le codage robuste et le codage conjoint source-canal pour la transmission vidéo sur l'Internet et sur des liens sans fils. Nous nous sommes également intéressés au codage scalable, y compris à granularité fine, pour apporter des solutions au contrôle de congestion dans des environnements de communication multipoints. Une étude de codage par descriptions multiples a démarré avec pour objectif de fournir des solutions de codage robuste avec une régulation de débit à granularité fine.*

6.3.1 Couplage source-réseau pour communications sur réseaux IP

Mots clés : Internet, réseaux IP, modèles de trafic, contrôle de congestion, qualité de service (QoS), vidéo.

Participants : Christine Guillemot, Xavier Hénocq, Reda Hosny, Fabrice Le Léanec, François Toutain, Jérôme Viéron.

Codage vidéo robuste : ces travaux sont menés en partie dans le cadre d'une CTI avec le Cnet. Ils visent un compromis optimal entre efficacité de compression et résistance aux pertes de paquets. En 1998, nous avons développé un algorithme adaptatif de choix de modes de codage intra ou inter, qui tient compte à la fois des caractéristiques du signal vidéo et des caractéristiques du réseau. Il permet ainsi d'optimiser le compromis entre l'efficacité de compression et la robustesse aux pertes de paquets. Dans la continuité de ces travaux, nous nous sommes attachés à mettre en oeuvre des modèles de performances débit-distorsion du codeur afin d'en accroître la faisabilité en temps réel. Ces modèles ont été testés et validés dans un codeur H.263 version 2. L'algorithme permet d'accroître fortement la robustesse aux pertes, en approchant la robustesse fournie par les approches classiques de rafraîchissement conditionnel, tout en conservant une efficacité de compression voisine de celle fournie par les schémas à base de prédiction temporelle, tels que les schémas MPEG ou H.26x (voir fig.8). L'estimation de la distorsion induite par le canal a également été optimisée par la mise en place d'un algorithme d'estimation de la propagation des pertes par estimation et compensation de mouvement.

Codage vidéo scalable : Les mécanismes de codage robuste ont été étendus à un mode de représentation scalable ou multi-niveaux des signaux vidéo. L'algorithme de sélection de modes selon les conditions de transmission, en plus de la distorsion due à la quantification et de la distorsion induite par le canal sur chaque niveau de scalabilité, tient compte des dépendances entre les niveaux de scalabilité et de l'effet de propagation des pertes inter-niveaux. Dans un



FIG. 8 – Codage vidéo robuste: (a) image de la séquence news codée par l'algorithme H.263 version 2, en présence de pertes sur le réseau; (b) image de la séquence "news" codée avec algorithme de choix de modes adaptatif aux caractéristiques du réseau

contexte de transmission multipoint, un contrôle de congestion réactif et précis peut exiger un nombre de niveaux de scalabilité élevé, 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 sélection conjointe des paramètres de quantification et des modes de codage, afin de minimiser la distorsion globale (source+canal).

Dans un contexte de flux pré-encodés, l'adaptation dynamique de débit de chaque niveau de scalabilité nécessite le développement de modèles de représentation scalable dits à *granularité fine*. L'un des problèmes posés est l'optimisation du compromis efficacité de compression et finesse de la granularité de la représentation scalable: les schémas à base de prédiction temporelle compensée en mouvement s'avèrent en effet être les plus performants en terme de compression mais ne permettent pas des ajustements fins du débit à des instants précis, en raison de la dépendance temporelle d'une image compressée à l'autre. Ils sont en outre très sensibles aux pertes et erreurs de transmission. Le développement d'un schéma de codage scalable à granularité fine est en cours.

Contrôle de congestion: Une première étude en 1998 avait mis en évidence un certain nombre de limitations de l'exploitation des modèles dits "TCP-friendly" (ou TCP-courtois) dans une approche de régulation de source vidéo. Ces limitations sont liées aux variations de trafics (en dents de scie, à l'instar de TCP) qui en résultent, ainsi qu'à l'absence de prédiction valable dans un environnement sans perte. Une première étude a alors conduit à la conception d'un algorithme de contrôle de congestion associant une prédiction de bande passante "TCP-courtoise" à un contrôle de délai d'aller-retour. Il a été montré au travers de simulations et d'expérimentations dans l'Internet qu'un tel contrôle se révèle plus adapté au pilotage de sources vidéo. Le mécanisme de régulation de débit développé permet en effet, dans un contexte de transmission avec pertes de paquets, de maintenir la qualité du signal vidéo constante, tout

en réduisant de manière significative le débit de la source. Ceci s'explique par une réduction de l'état de congestion du réseau, du fait de la diminution du débit de la source, qui se traduit par un effet bénéfique sur la qualité du signal reçu, pour un débit moindre. Cependant, les modèles de prédiction de bande passante de type "TCP-friendly" ne tiennent pas compte des contraintes de délais associées aux flux temps-réels. Ces contraintes de délais sont pourtant d'autant plus critiques que la technique de compression repose sur des mécanismes de prédiction, largement utilisés dans les schémas normalisés (H.26X, MPEG-X). Un algorithme permettant de transcrire les concepts de "TCP-courtoisie" et les contraintes de transmission en temps réel sous forme de contraintes fines de débit au niveau de la source vidéo a été développé et est en cours de test et de validation.

Contrôle de congestion multipoint : Plusieurs méthodes de contrôle de congestion multipoint, exploitant les fonctions de routage multicast du réseau ont été proposées dans la littérature. Ces approches font l'hypothèse d'un codage multi-niveaux ou codage scalable des signaux audio/vidéo, où chaque niveau correspond à un incrément de qualité en terme de définition temporelle (*scalabilité temporelle*), de rapport signal à bruit (*scalabilité SNR*) ou en terme de résolution spatiale (*scalabilité spatiale*). Les récepteurs viennent alors s'abonner ou se désabonner des groupes multicast correspondant aux différents niveaux de scalabilité. Cependant, ces techniques de contrôle guidées seulement par les récepteurs présentent un certain nombre de limitations: la première concerne la finesse d'adaptation du débit. Augmenter le nombre de niveaux de scalabilité, pour améliorer la granularité d'adaptation, va augmenter la charge induite par le mécanisme de contrôle de débit sans nécessairement résulter en un accroissement significatif de qualité. Nous avons alors étudié un mécanisme de contrôle hybride émetteur/récepteurs, travaillant par coopération entre ces entités, et reposant sur une représentation scalable des flux audio/vidéo. La régulation repose alors sur la maximisation d'une mesure de *goodput* tenant compte de modèles débit-distorsion de la source vidéo ainsi que de mécanismes de contrôle de pertes complémentaires.

Contrôle de Pertes, Codage conjoint source/canal : les techniques décrites ci-dessus - contrôle de congestion et régulation de débit associée, codage robuste à différents modèles de pertes - permettent d'adapter le débit des flux vidéo transmis aux caractéristiques du réseau, et ceci afin de contribuer à l'amélioration de la qualité de la transmission. La qualité apparente de la transmission peut aussi être améliorée en associant aux flux vidéo de l'information de redondance. La question alors posée est comment répartir au mieux la bande passante totale entre le flux vidéo et le flux contenant les données de redondance. Le projet a développé un mécanisme d'allocation de débit conjoint source-canal : connaissant la probabilité effective des pertes après décodage, la répartition de débit entre la source et la redondance (ou codage de canal) est réalisée de manière à minimiser la distorsion globale, définie comme la somme de la distorsion de quantification et de la distorsion induite par le canal, compte tenu des pertes résiduelles après correction. Cette optimisation globale de la chaîne de communication, incluant codage de source et codage de canal, avec une protection inégale des différentes portions de flux vidéo, permet un gain en terme de qualité (PSNR) du signal reconstruit. L'extension de l'algorithme à une représentation scalable des signaux vidéo associée à un contrôle de congestion hybride émetteur/récepteurs est en cours.

Différenciation de services : Une étude du modèle de différenciation de services, introduit à l'IETF, et de son impact sur des flux temps réel et en particulier sur la vidéo, a été démarré au travers d'une thèse. L'impact des paramètres des algorithmes mis en place au niveau des files d'attente (ex : Red - Random Early Detection) est étudié afin de dériver un choix optimum de ces paramètres d'un point de vue qualité de service des flux vidéo transmis d'une part et d'un point de vue d'un partage équitable de ressources entre des flux TCP et des flux vidéo temps-réel utilisant le protocole UDP d'autre part.

6.3.2 Codage conjoint source-canal pour communications sans fils

Mots clés : codage conjoint source-canal, codage par descriptions multiples, décodage conjoint, mobilité, modèles de canal.

Participants : Christine Guillemot, Francisco Garcia-Ugalde, Thomas Guionnet, Stéphane Pateux.

Les objectifs de forte compression conduisent à l'utilisation de codes statistiques à longueurs variables. Néanmoins, en plus de leur forte sensibilité au bruit, une difficulté supplémentaire est la dérivation analytique de leur performance débit-distorsion, en cas de corruption de bits. Nous avons développé une approche de protection inégale de sous-flux issus d'une décomposition par paquets d'ondelettes, basée sur une métrique de distorsion permettant d'estimer la distorsion induite par le canal dans un contexte d'utilisation de codes sources à longueurs variables. Ces métriques ainsi que les mécanismes d'allocation conjointe de débit de source et de débit de canal (codes RCPC - codes convolutionnels poinçonnés) sont en cours d'introduction dans un algorithme de codage H.263 version 2 pour répondre à des besoins de liaison vidéo numérique en robotique mobile. Nous avons, également, au travers d'un stage, étudié des transformations recouvrantes permettant de minimiser une erreur de reconstruction du signal en cas d'erreurs ou d'effacements. Ces travaux se prolongent aujourd'hui, d'une part par une étude postdoctorale de transformations $F : C^n \rightarrow C^m, m \geq n$, associées à des décompositions du signal sur des bases de fonctions $\Phi = \{\phi_k\}_{k=1}^m$ redondantes. L'approche - bases de fonction redondantes - s'apparente conceptuellement à un code de canal en bloc, de taux n/m , mis en oeuvre au travers de la transformation. Les paramètres n/m et les fonctions de base de la transformation peuvent être adaptés aux caractéristiques du canal. En parallèle, une thèse de doctorat se concentre sur la conception de quantificateurs à descriptions multiples, en essayant de combiner les propriétés liées à la description multiple à des caractéristiques de régulation de débit à granularité fine nécessaire pour des fonctions de contrôle de congestion ou d'adaptation dynamique des flux à la bande passante du réseau.

6.3.3 Régulation de débit pour la transmission d'images satellitaires sur canal à débit fixe

Participants : Delphine Le Guen, Stéphane Pateux, Claude Labit.

Dans le cadre de la transmission avec compression d'images satellitaires, nous avons mis en place un schéma de régulation locale basé sur une transformée en ondelettes de l'image. Ce schéma est basé sur une classification des coefficients d'ondelettes et offre la possibilité d'une

transmission progressive de l'information. Une telle technique permet d'améliorer la qualité locale dans l'image par l'octroi d'un débit plus important aux zones d'intérêt définies par la classification. Pour ce faire, nous avons développé une méthode intégrant plusieurs codeurs EZW distincts, s'appliquant chacun sur une des classes. De plus, en raison de problèmes de stockage à bord du satellite, nous ne pouvons traiter l'image entière et nous sommes donc dans l'obligation de la découper en fauchées de N lignes que nous codons successivement ; les codeurs de type EZW sont alors non seulement relatifs à une classe de coefficients mais aussi à une fauchée particulière. Un débit cible est alors défini pour chacune des classes sous la contrainte d'un débit global sur la fauchée, tout en privilégiant adaptativement les zones d'intérêt.

Les résultats obtenus (voir fig.9 et fig.10) montrent une amélioration significative de la qualité de l'image reconstruite par rapport à l'algorithme de régulation actuel de Spot5 pour un même taux de compression (généralement de l'ordre de 6). L'utilisation de la classification limite notamment les problèmes de rebond du signal situés dans les zones de contours proches de zones uniformes.

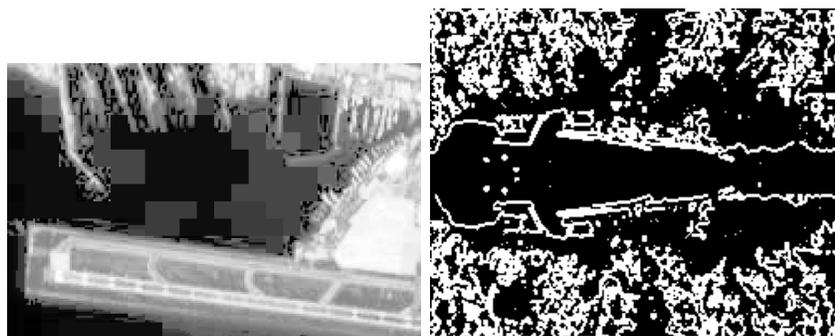


FIG. 9 – Image Gênes originale (2000x1600) et carte d'étiquettes

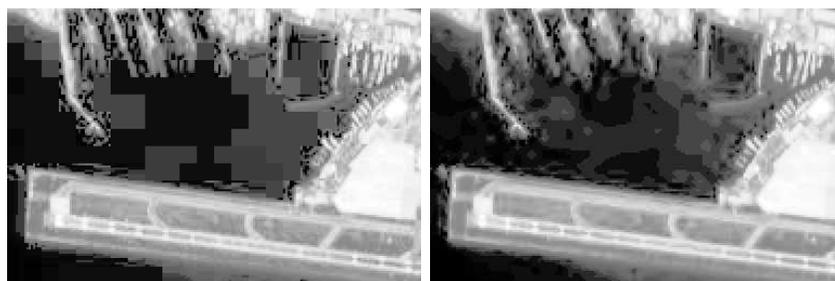


FIG. 10 – Technique de régulation SPOT5 et codage de type EZW classifié pour un taux de compression de 6

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Projet européen Esprit Nemesis

Participants : Henri Nicolas, Lionel Oisel.

Partenaires : Thomson CSF Optronique, Inria, HHI-Berlin, DMI-Tampere, ICS-Forth, UIB, Arbit, RTL.

Durée : 24 mois à partir du 1/9/98,

Réf. Inria : 196C4590031315006

Le monde de la post-production vidéo est actuellement très demandeur d'outils permettant de manipuler le contenu des images et de créer de nouvelles séquences vidéo en utilisant la notion d'objets vidéo. L'objectif principal du projet Nemesis est de développer et de commercialiser un ensemble d'outils algorithmiques semi-automatiques permettant une manipulation aisée des objets vidéos. De plus, une action de contribution à MPEG-4 est menée, portant sur une proposition d'extension de cette norme à la post-production vidéo. Pour cela le projet est organisé autour de trois partenaires techniques (Inria, HHI, et Forth), et de deux partenaires industriels (Thomson, Mikros), ce dernier étant spécialisé dans la post-production vidéo. Les fonctionnalités envisagées concernent la segmentation spatiale et spatiotemporelle des images, la création d'images panoramiques, le suivi d'objets 2D et 3D, la calibration interactive, le remplissage de textures. De plus, une étude complémentaire est menée concernant la représentation des données 2D et 3D, compatible à la fois avec la norme MPEG-4 ainsi qu'avec les modes actuels de représentation dans le monde industriel de la post-production vidéo. La contribution de TEMICS à ce projet est liée au développement d'outils de création d'images panoramiques, et à l'utilisation de ces images pour la manipulation des séquences d'origine. D'autre part Temics (H. Nicolas) assure la coordination du "workpackage 4" chargé des développements techniques du projet.

7.2 Projet européen ACTS-Comiqs: "Commerce thru MPEG-4 on the Internet with QoS"

Participants : Christine Guillemot, François Toutain, Jérôme Viéron.

Partenaires : CCETT, Philips-Lep, CSELT, université de Stuttgart, GIE Échangeur, DMI, Film Akademie, TraumWerk, Finsiel.

Durée : 24 mois à partir de mars 1998.

Réf. Inria : 1 98 C 107 00 31324 00 5

L'objectif du projet est le développement et la validation de nouvelles technologies dans le domaine de la communication multimédia sur Internet. Se plaçant au sein d'un processus de convergence d'applications multimédia - allant du domaine des jeux au domaine de l'ingénierie - les résultats techniques du projet Comiqs seront validés dans le domaine du commerce électronique. La valeur ajoutée et la contribution spécifique du projet Comiqs réside dans

l'implémentation et l'expérimentation des concepts suivants :

- intégration de tout type de média (audio, vidéo, naturels et contenus synthétiques 2D/3D)
- interactivité accrue avec les contenus - objets au sein de séquences animées ou de contenus sonores, objets 3D -
- streaming temps réel de tout type de média dans la scène, facilitant la navigation de l'utilisateur et permettant une meilleure adaptation du contenu présenté aux ressources disponibles du réseau
- positionnement de MPEG-4 dans un contexte Internet,
- contrôle de la Qualité de Service (QoS), d'une qualité "au-mieux" ou "best-effort" à une qualité prédite ou négociable.

Notre contribution porte sur l'adaptation des flux audiovisuels au transport sur Internet, le codage robuste des sources vidéos, les mécanismes de contrôle de congestion.

7.3 Projet RNRT-Osiam

Participants : Stéphane Pateux, Mustafa Sakalli.

Partenaires : Philips-Lep, Ireste, Irisa/Inria-Rennes, I3S, TIMC, Creatis, FT/Cnet.

Durée : 36 mois à partir de décembre 1998.

Réf. Inria : 2 98 C 562 00 31324011.

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.

7.4 Projet RNRT-COHRAINTE

Participant : Stéphane Pateux.

Partenaires : FT/Cnet, Philips-Lep, Irisa/Inria-Rennes, ENST-Paris.

Durée : 36 mois à partir de mars 1999.

Réf. Inria : 2 99 C 322 00 313 24 011.

Le projet exploratoire Cohrainte ("Codage hiérarchique et robuste de sources audiovisuelles et application à l'Internet") 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.

7.5 Projet RNRT-Visi

Participants : Christine Guillemot, Jérôme Viéron.

Partenaires : Thomson Multimédia, Inria, Edixia, Eurecom, FT/Cnet.

Durée : 24 mois à partir de juin 1999.

Réf. Inria : 2 99 A 461 00 OOOMC 011.

Le point fort du projet est l'appréhension, de manière conjointe, de différents verrous technologiques à savoir,

- la représentation hiérarchique des sources avec des mécanismes de régulation de débit appropriés,
- le contrôle de congestion et le contrôle d'erreur,
- les concepts architecturaux et systèmes liés à la répartition des sources entre plusieurs serveurs.

Il s'agit d'optimiser l'utilisation de la bande passante du réseau et plus généralement d'offrir des solutions au problème de la gestion de la qualité de service dans l'Internet. L'objectif est en outre de développer des solutions aussi pérennes que possibles, 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.

de scènes 3D

7.6 CTI-Cnet : Modélisation et représentation hiérarchique de scènes 3D

Participants : Luce Morin, Lionel Oisel.

Partenaires : INT-Evry (équipe Sim), Inria (Projet Syntim et Temics), Cnet

Durée : 36 mois à partir de mars 1996.

Réf. Inria : 1 96 A 227 00 000MC012.

Le thème général de cette collaboration était la modélisation et la représentation hiérarchique de scènes 3D pour des services multimédia. La contribution du projet a porté sur la reconstruction de vues par géométrie projective à partir de séquences vidéo monoculaires non calibrées.

sur réseaux hétérogènes

7.7 CTI-Cnet : Transmission multipoint de données multimédia sur réseaux hétérogènes

Participants : Christine Guillemot, Xavier Hénocq, Fabrice Le Léanec.

Partenaires : Inria, Cnet

Durée : 36 mois à partir de septembre 1998.

Réf. Inria : 1 98 C 391 00 31324 01 2.

Notre contribution porte sur les thèmes de recherche suivants:

- la recherche de modes de représentation scalable temporelle permettant de maintenir quelques modes de décorrélation temporelle des données video tout en assurant une communication sans propagations des pertes d'information sur un horizon temporel long,
- l'étude de modes de représentation scalable fréquentielle permettant une décorrélation optimale. Cette phase s'attache à l'étude des outils de type EZT ou paquets d'ondelettes,
- l'étude de mécanismes de codage robuste avec robustification non identique entre les sous-flux. Cette phase vise au développement de l'approche codage conjoint source-canal pour les réseaux utilisant les protocoles IP,
- 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.

7.8 Contrat Celar : Liaison vidéo numérique en robotique mobile

Participants : Thierry Fauconnier, Christine Guillemot.

Partenaires : Inria, Celar

Durée : 6 mois à partir de septembre 1999.

Réf. Inria : 1 99 C 459 00 31324 061.

Le problème posé consiste à transmettre un signal vidéo sur un réseau sans fil à partir d'un équipement d'acquisition mobile embarqué, pour une application de téléguidage d'un robot terrestre mobile. Le réseau support est donc 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 pertes et aux erreurs de transmission dans un environnement bruité. Il peut s'avérer nécessaire de mettre en oeuvre une coordination entre la source (codage source) et le mécanisme de transmission (codage canal), qui prenne en compte les paramètres principaux de la connexion à savoir le débit et les modèles d'erreurs.

7.9 Convention Cifre Microprocess

Participants : Henri Nicolas, Franck Denoual.

Durée : 36 mois à partir de mai 1997.

Thèse de : Franck Denoual

Réf. Inria : 1 97 C 531 00 31315 012

Cette collaboration avec la société Microprocess a pour objectif le développement de méthodes semi-automatiques de composition de scènes et leur implantation sur un noyau temps-réel. L'objectif pour la société Microprocess est d'étendre ses activités dans le domaine du temps-réel au monde de la compression vidéo MPEG et de la composition de scènes vidéo. Les

décodeurs MPEG 1,2 et 4 ont été implantés sur le noyau temps-réel Soft Kernel de la société Microprocess selon la méthodologie OMT. Une nouvelle méthode semi-automatique de mixage d'objets vidéo est également en cours de développement. Ce travail de développement logiciel a été réalisé en accord avec l'industriel.

7.10 Convention Cifre Thomson-TCO

Participant : Henri Nicolas.

Durée: 36 mois à partir de mai 1998.

Thèse de: Agnès De Simon.

Réf. Inria: 1 98 C 307 00 31324 012

Cette convention avec Thomson CSF Optronique concerne l'encadrement de la thèse d'Agnès Desimon sur le thème "Décomposition de séquences d'images par objets et applications". 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. Ce travail s'appuie sur les modes de représentation des objets vidéo définis par la norme MPEG-4, et consiste à développer de nouveaux outils permettant le mixage d'objets vidéo dans une séquence d'images incluant la gestion des problèmes de positionnement et de trajectoire.

7.11 Contrat Cnes

Participants : Claude Labit, Stéphane Pateux, Delphine Le Guen.

Durée: 36 mois à partir de octobre 1997.

Thèse de: Delphine Le Guen

Réf. Inria:

L'étude a pour objectif la définition de mécanismes de régulation de débit pour la transmission d'images satellitaires, en collaboration avec le Cnes (bourse Cnes pour le doctorat de D. Leguen) et Alcatel Aérospatiale Cannes. Cette recherche s'appuie sur un schéma de régulation réalisé antérieurement par le Cnes pour l'application Spot5. Celui-ci est basé sur une transformation DCT suivie d'une quantification scalaire avec un module de régulation intégré, activé par "fauchées" de 8 lignes d'images. Les résultats sont satisfaisants pour des taux de compression allant jusqu'à 2, mais la recherche de taux de compression plus élevés nous pousse à revoir les modules de compression et de régulation.

7.12 Convention Cifre Thomson Multimédia

Participants : Luce Morin, Philippe Robert.

Durée: 36 mois à partir de Octobre 1998.

Thèse de: Yannick Nicolas.

L'objectif de la 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,...).

7.13 Convention Cifre FT-Cnet

Participant : Christine Guillemot.

Durée : 36 mois à partir de février 1999.

Thèse de : Lionel Guivarch

L'objectif de la convention Cifre est de développer des approches de codage conjoint source-canal pour la transmission d'images dans des canaux de type radio-mobile. Pour cela trois principales voies d'études sont explorées:

- établissement de modèles de référence et de critères d'évaluation,
- recherche non contrainte de codeurs conjoints,
- recherche de solutions semi-compatibles.

À terme, elles prendront en compte des sources d'images animées et des canaux de type réseau local sans fil (HyperLAN ou Hiperaccess) :

8 Actions régionales, nationales et internationales

8.1 Actions régionales

Nous participons au groupe "Thélème" de réflexion sur le problème de la qualité de service dans la diffusion de compte-rendus médicaux. La consultation d'images par le praticien est un enjeu critique en terme de ressources réseaux et en terme de qualité de service. Les autres partenaires de ce groupe sont l'ENST-Bretagne, l'Ireste, les laboratoires Sim et Protimes de la faculté de médecine de Rennes, le laboratoire Latim du CHU de Brest.

8.2 Actions nationales

Le projet Temics participe au GDR-PRC Isis "Information, Signal et ImageS". Un partenariat sur le thème de la "télémédecine" et de la qualité de service en imagerie médicale coopérative a été mis en place avec des équipes de recherche Inria (Armor), Inserm (CHU Pitié-Salpêtrière), CNRS (Ireste-Nantes), le laboratoire Sim de l'université de médecine de Rennes et le Latim de l'université de Brest. Les principaux thèmes abordés sont les aspects QoS et applications de télémédecine. Ce partenariat a conduit en 1999 au dépôt d'une proposition de projet dans le cadre inter-régions Bretagne-Pays de Loire du contrat de plan Etat-Régions et d'une proposition d'action concertée incitative auprès du ministère.

8.3 Actions internationales

Le projet Temics participe à plusieurs projets et groupes de travail internationaux:

- une bourse individuelle post-doctorale européenne TMR ("Training and Mobility for Research") a été obtenue par J. Stauder (univ. Hannover, RFA) dans le cadre de l'appel à candidature 1997. Ce séjour post-doctoral de 18 mois au sein de notre projet a débuté en mai 1998,
- action Cost-211-3 et -4; participation de J. Stauder à cette action européenne sur les objets vidéo (segmentation, représentation, codage),
- participation à l'IETF - Internet Engineering Task Force -, et en particulier aux réunions des groupes AVT, RMTP et DiffServ,
- contribution à la normalisation MPEG-4 au travers de partenaires industriels, et contribution à la normalisation MPEG-7.

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

- programme France-Italie (participants: C. Labit, C. Guillemot): dans le cadre du programme franco-italien Galilée, nous menons une collaboration avec l'Université de Brescia (Prof. R. Leonardi). Le thème de cette action de collaboration est la compression d'images avec régions d'intérêt pour le télédiagnostic médical;
- programme France-Maroc (participants: C. Labit, L. morin, S. Pateux): Cette action intégrée franco-marocaine (responsable français: Luce Morin) est menée en collaboration avec l'université de Rabat, Maroc (Professeur D. Aboutajdine, assistant professeur de l'INPT-Rabat A. Tamtaoui). Dans le cadre de cette collaboration, le projet Temics a assuré deux formations en DEA de l'Université de Rabat et à l'INPT de Rabat sur les thèmes "vision par ordinateur" (L. Morin) et "codage conjoint source-canal" (F. Le Léanec);
- programme France-Mexique (participants: L. Morin, H. Nicolas, S. Pateux, C. Guillemot): Cette collaboration est menée avec l'Unam-Mexico (professeurs Garcia-Ugalde et Garcia-Garduno) dans le cadre d'une convention CNRS-Conacyt;
- collaboration France-Egypte (participant: C. Guillemot): Dans le cadre de cette collaboration, le projet Temics a assuré un module d'enseignement intitulé "communication vidéo" à l'ITI - Information Technology Institute - , Le Caire.

8.4 Visites, et invitations de chercheurs

Visite du professeur Y. Paker, QMW, Londres, du 13 septembre au 31 octobre 1999: cette visite a pour but de mettre en place une collaboration entre Temics et QMW sur le thème de la modélisation 3D et de la représentation scalable des informations associées pour des applications de diffusion et de communication multimédia.

9 Diffusion de résultats

9.1 Animation de la Communauté scientifique

- C. Guillemot est membre du comité de programme de la conférence Spie-VCIP 1999.
- C. Guillemot est membre du comité de programme de la conférence IEEE-Eurasip Workshop on Non linear Signal and Image Processing (NSIP 1999).
- C. Guillemot a participé à l'évaluation des propositions de projets de recherche européens IST, ainsi qu'à l'étude sur l'intégration et le management du programme IST.
- L. Morin, H. Nicolas et S. Pateux ont participé activement (présentations) aux réunions du GDR Isis.
- J. Stauder est membre de l'action européenne Cost 211quad

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, L Bonnaud, F. Garcia Ugalde : traitement d'images, vision 3D, mouvement, codage, compression d'images, cryptographie) ;
- Diic-TST, Ifsic, université de Rennes 1 (J. Stauder : réseaux de téléphone mobiles) ;
- 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 (L. Morin : analyse du mouvement 3D), (C. Guillemot : codage) ;
- Enic, Villeneuve-d'Ascq, (C. Labit : Objets vidéo et communication d'images numériques ; C. Guillemot: Vidéo sur IP) ;
- Ensar Rennes, option informatique (L. Morin : fondements du traitement d'image, morphologie mathématique) ;
- ENSTBR (C. Labit : Objets vidéo et communication d'images numériques, C. Guillemot : Les normes de codage vidéo de MPEG-1 à MPEG-4) ;
- ENS-Cachan (C. Labit : Objets vidéo et communication d'images numériques).

9.3 Autres actions de formation

- Stages effectués au sein du projet : Mickael Bourserie (DEA Stir, Rennes), Jérôme Viéron (DEA-Info, Rennes), Pascal Le Léanec (IIE, Paris), Fabrice Vergnenegre (DEA-Info), Brendan Catherine (ENSPS, Strasbourg), Bertrand Le Saux (Enserg), Bruno Gilles (DEA-Stir) ;

10 Bibliographie

Ouvrages et articles de référence de l'équipe

- [1] C. GUILLEMOT, E. CETIN AND R. ANSARI, *Nonrectangular Wavelet Representation of 2-D Signals. Application to Image Coding*, Eds Elsevier Press 1994, September 1994, p. 27–64.
- [2] 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.
- [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] E. NGUYEN, C. LABIT, « Selective coding by focus of attention: a new tool to achieve VLBR video coding (Invited Paper) », in : *Proceedings of Eusipco'96, Trieste, Italie, 1*, p. 431–434, September 1996.
- [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] A. TAMTAOUI, C. LABIT, « Constrained disparity and motion estimators for 3DTV image sequence coding », *Signal Processing: Image Communication* 4, 1991, p. 45–54.
- [8] G. TZIRITAS, C. LABIT, *Motion Analysis for Image Sequence Coding*, Elsevier Science, Book in Series “Advances in Image Communicatio”, 390 pages, 1994.
- [9] D. WANG, C. LABIT, J. RONSIN, « Morphological spatio-temporal simplification for video image segmentation », *IEEE Trans on Circuits and Systems for Video Technology* 7, 3, June 1997, p. 549–554.

Articles et chapitres de livre

- [10] C. GUILLEMOT ET T. TURLETTI, « Algorithmes de codage et de contrôle de transmission vidéo sur réseaux hétérogènes », chapitre dans livre - Systèmes Multimédia Communicants -, à paraître.
- [11] F. LE LÉANNEC, F. TOUTAIN ET C. GUILLEMOT, « Packet loss resilient MPEG-4 compliant video coding for the Internet », *Journal Image Communication, Numéro Spécial 'Real-time video over the Internet'*, 15, septembre 1999, p. 35–56.
- [12] H. NICOLAS ET F. JORDAN, « Interactive optimization of video compression algorithms on massively parallel computers », Chapitre 23 of the book Industrial Strength Parallel Computing.
- [13] E. IBN-ELHAJ, D. ABOUTAJDINE, S. PATEUX, L. MORIN, « HOS-based method of global motion estimation for noisy images sequences », *IEE Electronics Letters* 35, 16, août 1999.
- [14] P. RAULT ET C. GUILLEMOT, « Symmetric delay factorization: a generalized theory for paraunitary filter banks », *IEEE Trans. on Signal Processing* 47, 12, décembre 1999, p. 3315–3326.
- [15] J. STAUDER, R. MECH, J. OSTERMANN, « Detection of moving cast shadows for object segmentation », *IEEE Trans. on Multimedia* 1, 1, mars 1999, p. 65–76.
- [16] J. STAUDER, « Augmented reality with automatic illumination control incorporating ellipsoidal models », *IEEE Trans. on Multimedia* 1, 2, juin 1999, p. 136–145.

Communications à des congrès, colloques, etc.

- [17] C. GUILLEMOT, S. WESNER ET P. CHRIST, « Integrating MPEG-4 over the Internet », in : *Proceedings ECMAST'99*, mai 1999.

-
- [18] C. GUILLEMOT, V. LE GOFF ET F. GARCIA UGALDE, « Joint source and channel rate allocation with LVQ and Convolutional Codes », *in: Proceedings PCS'99, Picture Coding Symposium*, avril 1999.
- [19] F. LE LÉANNEC ET C. GUILLEMOT, « Error Resilient Video Transmission over the Internet », *in: Proceedings SPIE VCIP'99, Visual Communication and Image Processing*, janvier 1999.
- [20] F. TOUTAIN, F. LE LÉANNEC, C. GUILLEMOT, « Un procédé de contrôle de débit pour la transmission de vidéo MPEG-4 sur l'Internet », *in: Proceedings 17ème Colloque sur le Traitement du Signal et des Images, GRETSI'99*, septembre 1999.
- [21] F. TOUTAIN, « TCP-friendly point-to-point video-like source rate control », *in: Proceedings PVW'99, Packet Video Workshop*, avril 1999.
- [22] FRANCK DENOUAL ET H. NICOLAS, « Modifications artificielles de trajectoires d'objets vidéo », *in: Proceedings 17ème Colloque sur le Traitement du Signal et des Images, GRETSI'99, 1*, p. 195–198, septembre 1999.
- [23] F. GALPIN, L. MORIN, L. OISEL, « Génération d'un modèle VRML texturé à partir de séquences non calibrées », *in: Journées Francophones des Jeunes Chercheurs en Analyse d'Images et Perception Visuelle, ORASIS 99, Aussois, France*, p. 141–150, avril 1999.
- [24] F. GALPIN, L. MORIN, L. OISEL, « VRML model generation from uncalibrated sequences », *in: Proceedings Internation Workshop on Synthetic - Natural Hybrid Coding and Three Dimensional Imaging, Thira, Grèce*, p. 99–102, septembre 1999.
- [25] D. L. GUEN, C. LABIT, G. MOURY, D. LEBEDEFF, « Régulation de débit et adaptation locale de qualité pour des systèmes de compression embarqués sur satellites », *in: Proceedings 17ème Colloque sur le Traitement du Signal et des Images, GRETSI'99, 1*, p. 47–50, Vannes, septembre 1999.
- [26] H. NICOLAS, « Critère optimal pour la création de séquences d'images panoramiques », *in: Proceedings 17ème Colloque sur le Traitement du Signal et des Images, GRETSI'99, 2*, p. 503–506, septembre 1999.
- [27] H. NICOLAS, « Nouveaux critères analytiques pour la création de séquences d'images panoramiques », *in: Proceedings colloque TAIMA*, Hammamet, Tunisie, mars 1999.
- [28] H. NICOLAS, « Optimal criterion for dynamic mosaicking », *in: Proceedings ICIP'99, Osaka, Kobe, Japon*, octobre 1999.
- [29] E. IBN-ELHAJ, D. ABOUTAJDINE, S. PATEUX, L. MORIN, « Méthode région-réursive d'estimation de mouvement des séquences d'images bruitées », *in: Proceedings 17ème Colloque sur le Traitement du Signal et des Images, GRETSI'99, 3*, p. 171–174, Vannes, septembre 1999.
- [30] J. MOTSCH ET H. NICOLAS, « Object-based motion classification », *in: Proceedings ICIP'99, Kobe, Japon*, octobre 1999.
- [31] J. MOTSCH ET H. NICOLAS, « VOP classification based on motion analysis », *in: Proceedings Workshop WIAMIS'99*, p. 89–92, juin 1999.
- [32] J. STAUDER ET H. NICOLAS, « Motion-based video indexing evaluating object shading », *in: Proceedings ICIP'99, Osaka, Kobe, Japon*, octobre 1999.
- [33] JEAN MOTSCH ET H. NICOLAS, « Classification du mouvement des objets », *in: Proceedings 17ème Colloque sur le Traitement du Signal et des Images, GRETSI'99, 1*, p. 183–186, septembre 1999.
- [34] G. MARQUANT, S. PATEUX, C. LABIT, « Codage d'images fixes par maillages », *in: CORESA '99*, Sophia-Antipolis, juin 1999.
- [35] L. OISEL, L. MORIN, E. MÉMIN, C. LABIT, « Reconstruction 3D de scènes complexes par maillage de cartes de disparité », *in: Proceedings 17ème Colloque sur le Traitement du Signal et des Images, GRETSI'99, 2*, p. 391–394, septembre 1999.

- [36] S. PATEUX, C. LABIT, « Optimisation débit-distorsion d'un schéma de codage vidéo orienté-régions », *in : Proceedings 17ème Colloque sur le Traitement du Signal et des Images, GRETSI'99*, 3, p. 849–852, Vannes, septembre 1999.
- [37] S. PATEUX, C. LABIT, « Rate-distortion optimized region-based video coder », *in : Proceedings ICIP'99*, Kobe, Japon, octobre 1999.

Rapports de recherche et publications internes

- [38] C. GUILLEMOT, P. CHRIST AND S. WESNER, « RTP Payload Format for MPEG-4 ES with Scalable and Flexible Error Resilience », *rapport de recherche*, Internet-draft, juillet 1999.