

Rapport mi-parcours DEMI-TON Description multimodale pour la structuration automatique des flux de télévision

1 Liste des équipes impliquées

TEXMEX : techniques d'exploitation des données multimédias, IRISA, Rennes

METISS : Modélisation et expérimentation pour le traitement des informations et des signaux sonores, IRISA, Rennes

DCA : Description des contenus audiovisuels, INA, Bry sur Marne

2 Liste des participants au 1/4/05

TEXMEX

- Gros, Patrick, directeur de recherche INRIA, 40 %
- Sébillot, Pascale, professeur INSA de Rennes, 35 %
- Kijak, Ewa, maître de conférences université de Rennes 1 depuis le 1/9/2006, 10 %.
- Naturel, Xavier, doctorant, contrat de recherche doctorale INRIA, 50 %, début de la thèse 1/2/2004.
- Huet, Stéphane, doctorant, allocataire MENESR, 50 %, début de la thèse 1/10/2004
- Dupuis, Arnaud, post-doctorant payé par l'ACI du 1/5/2005 au 30/11/2005, 100 %, ingénieur expert INRIA payé sur un contrat du FNADT depuis le 1/12/2005, 100 %.
- Dufouil, Cédric, Ingénieur associé INRIA (payé sur subvention INRIA), 100 %.

METISS

- Gravier, Guillaume, chargé de recherche CNRS, 50 %
- Moraru, Daniel, post-doctorant payé par l'ACI du 1/5/2005 au 31/1/2006, 100 %.

DCA

- Brunie, Vincent, chercheur INA, 30 %, jusqu'à fin avril 2006.
- Carrive, Jean, chercheur INA, 30 %
- Vinet, Laurent, chercheur INA, 30 %
- Balin, Fabrice, ingénieur INA, 30 %
- Poli, Jean-Philippe, doctorant CIFRE INA, 100 %, début de la thèse 4/1/2004.

3 Changements significatifs intervenus dans le projet

- V. Brunie qui était responsable du projet pour l'INA a quitté l'INA. Il a été remplacé par J. Carrive qui était déjà impliqué dans le projet.

4 Résumé des principales avancées

4.1 Structuration de flux de télévision

Un des buts principaux du projet était de décrire des flux de télévision de longue durée (au minimum plusieurs semaines). Deux approches ont été étudiées.

Approche prédictive. La thèse de J.P. Poli à l'INA a adopté un point de vue top-down, en partant de l'observation des guides de programmes, ceux, de type prévisionnel, publiés dans la presse spécialisée et ceux établis par l'INA à partir de l'observation du flux réel. En modélisant les grilles de programmes d'une chaîne et en entraînant le modèle sur les grilles des années passées, il est possible de déduire tous les enchaînements possibles d'émissions, y compris celles qui ne sont pas annoncées dans les guides. Le modèle est constitué d'un modèle de Markov contextualisé, pour les transitions d'un programme à l'autre, et d'un arbre de régression, utilisé pour prédire la durée d'une émission. Ce modèle permet de prédire à quelle sorte de transition il faut s'attendre, afin de savoir quel objet doit être détecté (un jingle de publicité, un générique d'émission) et surtout à quel endroit du flux il doit être recherché. Ainsi les détections sont effectuées localement dans le flux et il a été montré expérimentalement que dans le pire des cas, moins de la moitié du flux était nécessaire pour le structurer. Ce travail a été réalisé en utilisant les guides de programme correspondant à un flux de plus d'un an de télévision.

Approche descriptive. La thèse de X. Naturel à l'IRISA a pris le point de vue inverse : à partir d'un guide programme prédictif (issu de la presse spécialisée) et du flux lui-même, on a cherché à identifier les différents programmes

et inter-programmes dans le flux en détectant leurs bornes et en étiquetant les segments ainsi délimités. La méthode utilisée consiste à tout d'abord détecter les inter-programmes à l'aide de marqueurs spécifiques et d'une base de tous les inter-programmes déjà détectés, puis d'aligner les segments de programmes avec ceux du guide de programme par DTW (Dynamic Time Warping).

Perspectives. Elles sont de deux types : tout d'abord la fusion entre les deux approches précédentes, mais aussi l'étude de la robustesse de ces méthodes sur des chaînes de télévision différentes et plus nombreuses.

4.2 Traitement de la bande sonore

Deux aspects ont été étudiés.

Analyse du flux de parole télévisuelle. Dans un premier temps, nous avons complété et amélioré la plate-forme IRENE pré-existante, adaptée à la transcription des informations diffusées à la radio. Nous avons d'une part ajouté un module permettant la structuration du flux selon le locuteur et étudié l'apport du suivi de locuteur pour améliorer la segmentation. D'autre part, nous avons exploité cette segmentation pour améliorer la transcription par l'ajout d'un module d'adaptation au locuteur. Cette nouvelle version de IRENE a ensuite été évaluée sur le flux télévisuel, mettant ainsi en évidence le manque de robustesse des modèles appris sur la radio devant la qualité sonore moindre de la télévision. Le travail s'est ensuite orienté sur l'utilisation de points d'ancrage macrophonétique pour améliorer cette robustesse et s'affranchir de la nécessité de disposer d'un grand corpus annoté de parole télévisuelle pour apprendre des modèles.

Lien entre RAP et TAL. Dans un premier travail, mené autour de la thèse de S. Huet, le lien entre reconnaissance automatique de la parole (RAP) et traitement automatique des langues (TAL) a été étudié. Le but était double : voir comment les techniques de TAL peuvent améliorer la transcription de la parole, et évaluer comment les techniques de TAL peuvent traiter la parole transcrite. Le premier aspect a donné lieu à un travail intéressant sur l'utilisation d'un tagger statistique PoS (Part of Speech = parties du discours) pour corriger les erreurs de la transcription : de nombreuses variantes morphologiques d'un même mot sont en effet homophones et ne peuvent être distinguées par un système de RAP classique. La prise en compte des caractéristiques PoS (genre, nombre, type du mot : nom, verbe...) permet une désambiguïsation dans de nombreux cas.

5 Réalisations obtenues dans le cadre du projet

Au delà des échanges purement scientifiques, le projet DEMI-TON a favorisé de nombreux échanges technologiques entre les deux équipes. Initialement, chaque entité avait amorcé le développement d'une plate-forme informatique adaptée au

traitement de documents multimédias. Dans le cadre de DEMI-TON, les deux équipes ont déterminé des points de convergence indispensables au partage et à la portabilité d'algorithmes de traitement entre les deux plates-formes.

Projet de plate-forme multimédia Les activités de recherche sur la structuration automatique de flux télévisés nécessitent de disposer de corpus audiovisuels significatifs représentant plusieurs semaines de programmes. Sachant que 24 heures de vidéo correspondent approximativement à 40 Go de données, les enregistrements de plusieurs semaines peuvent alors occuper des espaces disque de plusieurs To. À ces informations s'ajoutent des données de description (grille des programmes, résumé...) ainsi que les résultats des outils de traitement (mouvement de caméra, suivi de visages...).

Par ailleurs, les outils d'analyse de vidéos étant souvent gourmands en ressources processeur, le traitement de ces gros volumes de données requière de disposer de puissances de calcul dimensionnées en conséquence. A ces problèmes de stockage et de calcul, s'ajoute la protection des droits d'auteurs sur les données audiovisuelles réglementant leur diffusion ainsi que leur duplication. Enfin, les natures hétérogènes des informations à traiter (image, son, texte) pose de nombreux problèmes de synchronisation et de précision d'accès aux données. Typiquement, l'accès aléatoire à une image (rechercher une image précise sans lire les images précédentes) au sein d'un flux vidéo MPEG, pourtant primordiale dans un processus de développement d'algorithmes d'analyse automatique de vidéos, peut rapidement s'avérer être complexe à mettre en oeuvre. Pour répondre à toutes ces contraintes techniques (stockage, limitation des droits d'accès, optimisation des calculs...), il apparaît donc nécessaire de développer une infrastructure informatique adaptée au traitement de données audiovisuelles.

Avancement du projet. Depuis mai 2005, une plate-forme de traitement de vidéos a été développée en collaboration avec l'INA. Cette plate-forme est d'une part destinée à favoriser les échanges d'outils de traitement entre l'INA et l'IRISA, et d'autre part, conçue pour répondre à de nombreux problèmes récurrents liés aux traitements de vidéos.

Une première phase de développement a été effectuée à l'IRISA de mai à novembre 2005 et a consisté à étudier et à développer un serveur centralisant des corpus audiovisuels pouvant représenter plusieurs semaines de programmes télévisés. Ce serveur devait permettre de programmer des acquisitions de séquences télévisées, de les diffuser en « streaming », de gérer les vidéos propres à chaque utilisateur et enfin de protéger les corpus en contrôlant les accès par le biais de login et mot de passe. L'accès à la plate-forme, via un outil baptisé « DIVA manager », s'effectue par le biais d'une interface web fonctionnant sous de nombreux systèmes d'exploitation (Windows, Linux et MacOSx).

Une seconde phase de développement a débuté en septembre 2005 et a consisté à développer une application client/serveur offrant un accès précis aux images et aux sons extraits de contenus audiovisuels. Cette application a été conçue

dans l'objectif de simplifier le développement d'outils de traitement vidéo et propose aux utilisateurs de faire abstraction des nombreux problèmes liés au décodage de séquences audiovisuelles. L'étude et le développement de l'application « DIVA client/server » a contribué à lever de nombreux verrous techniques liés à la précision d'accès ou encore à la synchronisation des données.

L'ensemble de cette solution logicielle, intégrant « DIVA manager » et « DIVA client/server » a été baptisée « DIVA solution ».

Valorisation des activités de TEXMEX associées à la plate-forme.

L'application « DIVA client/server » (distant video access), développée à l'IRISA, a fait l'objet d'un dépôt de code au près de l'APP (agence pour la protection des programmes), elle est référencée sous le numéro : IDDN.FR.001.320006.000.S.P.2006.000.40000. Parallèlement, un article décrivant le système ainsi que les problèmes techniques résolus par le logiciel a été publié lors de la semaine du document numérique (SDN'06) dans le cadre de l'atelier SIAV (Systèmes d'Information AudioVisuels).

6 Réunions et Conférences organisées dans le cadre du projet

- Début du projet : 1er avril 2005
- Réunion plénière du projet : 11 octobre 2005, Rennes
- Séminaire technique : du 24 au 28 octobre 2005, Bry sur Marne
- Réunion plénière du projet : 30 janvier 2006, Paris
- Réunion plénière du projet : 6 juin 2006, Rennes
- Séminaire technique : 3 et 4 juillet 2006, Bry sur Marne
- Réunion plénière du projet : 28 novembre 2006, Rennes

7 Soutiens obtenus en liaison avec ce projet

7.1 Postes chercheurs

Xavier Naturel : doctorant, contrat de recherche doctorale INRIA, début de la thèse 1/2/2004.

Jean-Philippe Poli : doctorant, convention CIFRE, début de la thèse 4/1/2004.

7.2 Postes ingénieurs

Arnaud Dupuis : après son séjour post-doctoral payé sur les crédits obtenus dans le cadre du projet DEMI-TON, A. Dupuis a été embauché comme ingénieur de niveau post-doctoral par l'INRIA grâce à un financement octroyé par le FNADT (Fonds national d'aménagement et de développement du territoire). Cette embauche a été effective le 1er décembre 2005 pour une durée de 3 ans.

Cédric Dufouil : a été embauché par l'INRIA (sur sa dotation budgétaire) pour deux ans le 1er septembre 2005 comme ingénieur en CDD.

7.3 Contrats nationaux

Semim@ges : 2007 - 2009, contrat ANR RIAM, montant en cours de négociation. Participation de l'IRISA.

7.4 Contrats européens

MUSCLE : réseau d'excellence du 6e PCRD Multimedia Understanding Through Semantics, Computation and Learning, 2004 - 2008. Dans ce cadre, nous menons des travaux sur la structuration multimodale des retransmissions sportives à la télévision et sur les modèles stochastiques qui permettent une telle structuration. Par rapport à nos activités dans DEMI-TON, il s'agit ici d'une structuration plus fine au niveau d'une seule émission. Il y a donc une forte complémentarité.

7.5 Contrats internationaux hors CEE

QUAERO : 2007 - 2011, contrat A2I, montant en cours de négociation, participation des 3 équipes de DEMI-TON.

7.6 Contacts internationaux dans le cadre de ce projet

Stéphane Marchand-Maillet : Centre universitaire d'informatique, université de Genève, Suisse.

Ichiro Ide : Université de Nagoya, Japon.

7.7 Soutiens financiers

- Le FNADT nous a accordé 150 kEUR qui nous ont permis d'embaucher Arnaud Dupuis comme ingénieur en CDD pour une durée de 3 ans.
- L'université de Rennes 1 nous a accordé 28 kEUR de BQR pour participer à l'équipement de la plate-forme.

- L'INSA de Rennes nous a accordé 13,5 kEUR de BQR pour tester l'emploi d'une baie RAID au sein d'un cluster de la grille GRID5000.
- l'INRIA nous a accordé 130 kEUR pour acheter le matériel de la version v3 de la plate-forme.

8 Publications obtenues dans le cadre du projet

Conférences internationales

- Jean-Philippe Poli "Predicting program guides for video structuring", in: Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence, pp. 407-411, Hong-Kong, Chine, novembre 2005.
- Jean-Philippe Poli, Jean Carrive "Video Stream Structuring and Annotation Using Electronic Program Guides", in: 5th International Workshop on Knowledge Markup and Semantic Annotation, pp. 137-141, Galway, Irlande, novembre 2005.
- Xavier Naturel, Patrick Gros. A Fast Shot Matching Strategy for detecting duplicate sequences in a television stream. CVDB'05: Proceedings of the 2nd ACM SIGMOD International Workshop on Computer Vision meets DataBases, Baltimore, USA, juin 2005.
- Stéphane Huet, Guillaume Gravier, Pascale Sébillot. Are Morphosyntactic Taggers Suitable to Improve Automatic Transcription?. Proc. of Text, Speech and Dialogue (TSD), Lecture Notes in Computer Science, Volume 4188/2006, pages 391-398, Brno, Tchéquie, Septembre 2006.
- Xavier Naturel, Guillaume Gravier, P. Gros. Fast Structuring of Large Television Streams using Program Guides. 4th International Workshop on Adaptive Multimedia Retrieval (AMR), Genève, Suisse, juillet 2006.
- Daniel Moraru, Mathieu Ben, Guillaume Gravier. Experiments on speaker tracking and segmentation in radio broadcast news. In *European Conference on Speech Communication and Technology – Interspeech*, 2005.

Journaux nationaux

- Patrick Gros. Description et indexation automatiques des documents multimédias : du fantôme à la réalité. *Documentaliste - Sciences de l'information*, 42(6):383-391, décembre 2005.
- B. Bachimont, Patrick Gros. Recherche : des défis scientifiques. Les nouveaux dossiers de l'audiovisuel, numéro spécial : Internet : quelle place pour la vidéo ?, (9):28-30, mars 2006.

Conférences nationales ou francophones

- Jean-Philippe Poli, Jean Carrive "Proposition d'une architecture pour un système de structuration automatique de flux audiovisuels", in: Actes de la Conférence sur la COmpression et la REprésentation des Signaux Audiovisuels 2005, pp. 49-53, Rennes, France, novembre 2005.
- Xavier Naturel, Guillaume Gravier, Patrick Gros. Étiquetage Automatique de Programmes de Télévision. CORESA'05 Compression et représentation des signaux audiovisuels, Rennes, France, novembre 2005.
- Stéphane Huet, Guillaume Gravier, Pascale Sébillot. Peut-on utiliser les étiqueteurs morphosyntaxiques pour améliorer la transcription automatique ?. Actes des 26èmes Journées d'Études sur la Parole (JEP), Dinard, France, juin 2006.
- Daniel Moraru, Guillaume Gravier. Ancres macrophonétiques pour la transcription automatique. Actes des 26èmes Journées d'Études sur la Parole (JEP), Dinard, France, juin 2006.
- A. Dupuis et C. Dufouil, « Une solution client/serveur pour l'analyse de corpus audiovisuels », SIAV'06 (Systèmes d'Information AudioVisuels), Fribourg Suisse, 18-22 septembre 2006.

Rapports de recherche

- Stéphane Huet, Pascale Sébillot, Guillaume Gravier. Introduction de connaissances linguistiques en reconnaissance de la parole : un état de l'art. Rapport de Recherche IRISA, No1804, mai 2006.
- Xavier Naturel, Patrick Gros. Detecting Repeats for Video Structuring. Rapport de Recherche IRISA, No 1790, mars 2006.