



Présentation des projets financés au titre de l'édition 2010 du
Programme « Contenus et interactions »
Défi Multimédia : REconnaissance de PERsonnes
dans des Émissions audiovisuelles (REPERE)

ACRONYME et titre du projet	Page
PERCOL – Reconnaissance de personnes dans des contenus audiovisuels	2
QCOMPERE – Consortium Quaero pour la Reconnaissance Multimodale des Personnes	4
SODA – reconnaiSsance de persOnnes pour Débats et les journAux télévisés	6

Programme REPERE - CONTINT

Edition 2010

Titre du projet	PERCOL – Reconnaissance de personnes dans des contenus audiovisuels
Résumé	<p>Le but du projet PERCOL consiste à participer au défi REPERE en agrégeant au sein d'un même consortium des partenaires issus du milieu industriel (Orange Labs) et académique (LIA, LIF, LIFL). La complémentarité de ces partenaires permet de structurer le projet PERCOL en trois axes :</p> <p>1) Utilisation de technologies existantes pour couvrir les différentes modalités prévues dans le défi REPERE. Cet axe prévoit l'utilisation de « briques » déjà existantes chez les différents partenaires et permettant, en outre, de répondre aux tâches suivantes :</p> <ul style="list-style-type: none">- Reconnaissance et segmentation en locuteur d'un flux audio.- Transcription automatique de la parole.- Localisation et reconnaissance de caractères incrustés dans des images.- Détection d'entités nommées dans du texte et de l'audio.- Reconnaissance automatique de visage.- Analyse de comportements à partir de documents vidéo. <p>2) Développement d'axes de recherche novateurs, s'inscrivant dans le cadre de REPERE, et permettant à la fois le croisement de « cultures » scientifiques différentes (audio, texte, vidéo) et l'avancement de l'état de l'art concernant la tâche principale du défi. Les deux principaux axes de recherche qui seront étudiés sont :</p> <ul style="list-style-type: none">- Algorithmes de fusion de descripteurs correspondant à différentes modalités (audio, texte, vidéo) pour la détection de personnes dans des flux audio.- Amélioration des méthodes d'extraction des différents descripteurs grâce à l'introduction d'informations provenant d'une autre modalité. Par exemple on pourrait utiliser des informations provenant de l'analyse de scènes vidéo pour aider à déterminer le nombre de locuteurs d'un segment et ainsi améliorer la phase de segmentation en locuteurs, ce qui conduira aussi à améliorer la reconnaissance de la parole par une meilleure adaptation au locuteur des modèles acoustiques.

3) Introduction de la notion d'usage dans la mise au point et l'optimisation des systèmes de détection de personnes qui seront développés pour participer aux évaluations du défi REPERE. Cet aspect est particulièrement important pour permettre de développer des systèmes correspondant à des cadres applicatifs réalistes. Cet axe de travail est rendu possible par la participation au consortium du groupe traitant de l'ergonomie et des usages de la plateforme d'indexation et de recherche vidéo d'Orange Labs. La plateforme de démonstration d'Orange Labs pourra également être utilisée pour intégrer les résultats du projet qui auront été développés sous forme de web services.

Partenaires

Laboratoire d'Informatique Fondamentale de Marseille (LIF)
Université d'Avignon et des Pays de Vaucluse (UAPV)
Laboratoire d'Informatique Fondamentale de Lille (LIFL)
France Télécom

Coordinateur

Frédéric Béchet - LIF
Frederic.Bechet@lif.univ-mrs.fr

Aide de l'ANR

397 800 €

Début et durée

01/11/2010 - 36 mois

Label pôle

Images & Réseaux
SCS (Solutions Communicantes Sécurisées)
Industries du Commerce

Programme REPERE - CONTINT

Edition 2010

Titre du projet	QCOMPERE – Consortium Quaero pour la Reconnaissance Multimodale des Personnes
Résumé	<p>Le projet QCOMPERE a pour objectif de constituer un consortium pour participer au défi multimédia REPERE. Le programme QUAERO, démarré en mai 2008, est composé d'organismes de recherche français et allemands. L'objectif du projet QUAERO « Core Technology Cluster » est d'améliorer l'état de l'art en structuration et indexation automatique de documents multimédia. Plusieurs de ses tâches sont directement liées à l'appel à projets REPERE, et chacun de ses partenaires, dans son domaine respectif, a montré une expertise technologique du plus haut niveau. La participation à REPERE est donc une extension naturelle du travail planifié dans QUAERO.</p> <p>Le projet sera organisé en six tâches, avec quatre tâches répondant chacune à un problème mono-modal :</p> <ul style="list-style-type: none">- qui est visible ?- qui parle ?- quel nom est écrit à l'écran ?- et quel nom a été prononcé ? <p>Une cinquième tâche s'occupe de la fusion multimodale et une sixième tâche fournira la coordination technologique et des ressources sous forme de données annotées nécessaires au déroulement du projet.</p> <p>L'objectif principal est la participation au défi REPERE, et les livrables majeurs sont logiquement les soumissions aux deux évaluations officielles de REPERE pour chacune des modalités. Les autres retombées attendues sont la collaboration scientifique et les publications scientifiques. La collaboration avec le partenaire allemand, non financé par le projet, sera principalement supportée par des missions d'échanges.</p>
Partenaires	<p>CNRS – Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) Centre de Recherche INRIA Grenoble Rhône-Alpes Laboratoire d'Informatique de Grenoble (LIG) YACAST Vocapia Research Groupe de Recherche en informatique, image, automatique et instrumentation de Caen (GREYC)</p>

Coordinateur Claude Barras - LIMSI CNRS
barras@limsi.fr

Aide de l'ANR 400 000 €

Début et durée 01/11/2010 - 36 mois

Label pôle -

Programme REPERE - CONTINT

Edition 2010

Titre du projet	SODA – reconnaissAnce de persOnnes pour Débats et les journAux télévisés
Résumé	<p>L'appel à projets REPERE est une campagne d'évaluation des systèmes de reconnaissance des personnes présentes dans des émissions audiovisuelles en français. Les systèmes concourants devront s'appuyer sur les différentes sources d'information présentes dans les émissions pour déterminer qui apparaît à l'image, qui parle, quels sont les noms et prénoms affichés à l'écran ou prononcés et à qui ils correspondent.</p> <p>Ces objectifs nécessitent des compétences variées. La reconnaissance des locuteurs et des visages nécessite une expérience solide en traitement automatique de l'audio, de la vidéo et en fusion de données. La maîtrise de la reconnaissance de parole, de la reconnaissance de caractères et du traitement de la langue naturelle permet d'extraire et d'associer correctement les noms de ces personnes.</p> <p>Les laboratoires impliqués dans ce projet de recherche sont au nombre de deux : le Laboratoire Informatique de l'Université du Maine (LIUM) qui dirigera le projet, et l'institut de recherche suisse IDIAP. Les compétences des partenaires permettent de couvrir tous les thèmes du défi. L'équipe Parole du LIUM a développé depuis 2004 un système de segmentation et de regroupement en locuteurs (SRL) performant qui a été classé 2^e lors de la campagne d'évaluation ESTER en 2005 et 1^{er} lors de la campagne d'évaluation ESTER 2 en 2008. Le LIUM travaille également depuis 2006 sur l'identification des locuteurs à partir de la transcription de la parole et des noms prononcés dans ces enregistrements. Le système d'identification s'appuie sur le système de reconnaissance de la parole à grand vocabulaire du LIUM développé depuis 2004. De son côté, l'IDIAP a développé depuis de nombreuses années des compétences en traitement automatique de l'audio et de la vidéo. Dans le cadre du projet, l'IDIAP s'appuiera principalement sur ses travaux en détection, suivi et reconnaissance des personnes et en reconnaissance de caractères. L'IDIAP a participé régulièrement aux évaluations NIST en reconnaissance de locuteurs. Il a participé et conduit la tâche de suivi et de reconnaissance de l'orientation de la tête dans l'évaluation</p>

internationale CLEAR (Classification of Events, Activities, and Relationships) en 2006 et 2007. Il conduit actuellement une campagne d'évaluation de reconnaissance de visages et de locuteurs à ICPR 2010 (International Conference on Pattern Recognition).

Le projet bénéficiera des résultats précédemment obtenus par les partenaires, mais nécessitera des efforts d'intégration afin de construire un système complet permettant de suivre les intervenants principaux ainsi que de les nommer tout au long de l'émission. Du point de vue de la recherche, la combinaison d'informations en provenance des différentes sources (signal acoustique, images, parole et texte) fédèrera nos travaux.

Par ailleurs, une originalité du projet portera sur la reconnaissance des différents rôles (présentateur, journaliste ou modérateur, invité) des personnes, de leurs relations et leur interaction (qui s'adresse à qui par exemple), ainsi que de l'exploitation de la reconnaissance de ces rôles pour améliorer les performances de la segmentation ou de l'association entre une voix et un visage. Par exemple, les interactions pourront aider la détection de la personne s'exprimant parmi les personnes à l'image, tandis que les rôles pourront faciliter l'identification des locuteurs pour cibler les zones susceptibles de contenir de l'information utile.

Partenaires

Laboratoire d'Informatique de l'Université du Maine (LIUM)
Idiap Research Institute

Coordinateur

Sylvain Meignier - LIUM
sylvain.meignier@lium.univ-lemans.fr

Aide de l'ANR

432 380 €

Début et durée

01/11/2010 - 36 mois

Label pôle

Images & Réseaux