Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 08-19 juin 2020

Projet AMIS : résumé et traduction automatique de vidéos

Mohamed Amine Menacer Dominique Fohr Denis Jouvet Karima Abidi
David Langlois Kamel Smaïli
Université de Lorraine, CNRS, LORIA, F-54000 Nancy, France
prenom.nom@loria.fr

RÉSUMÉ

La démonstration de résumé et de traduction automatique de vidéos résulte de nos travaux dans le projet AMIS. L'objectif du projet était d'aider un voyageur à comprendre les nouvelles dans un pays étranger. Pour cela, le projet propose de résumer et traduire automatiquement une vidéo en langue étrangère (ici, l'arabe). Un autre objectif du projet était aussi de comparer les opinions et sentiments exprimés dans plusieurs vidéos comparables. La démonstration porte sur l'aspect résumé, transcription et traduction. Les exemples montrés permettront de comprendre et mesurer qualitativement les résultats du projet.

ABSTRACT

AMIS project : automatic summarization and translation of videos

The demonstration of video summarization and machine translation is the result of our work in the AMIS project. The project aimed at helping a traveller to understand news in a foreign country. For that, the solution was to provide summaries and translations of videos given in a foreign language (here Arabic). Another goal of the project was to compare sentiments and opinions expressed in comparable videos. to compare in terms of sentiments and opinions comparable videos. During the demo, we will present several videos in Arabic, and their translation in English. The given examples will be representative of the results of the project.

MOTS-CLÉS : vidéo, résumé automatique, traduction automatique, arabe, anglais.

KEYWORDS: video, automatic summarization, machine translation, Arabic, English.

1 Introduction

Imaginons un voyageur dans une ville à l'étranger. Un événement local survient, qui fait la une des journaux télévisés. Le voyageur passe d'une chaîne à une autre, mais ne comprend que partiellement ce qu'il se passe. Pourtant, cela peut avoir un impact fort sur son séjour (fermeture d'aéroport, limitation de circulation). Le voyageur a donc besoin, depuis son hôtel, d'avoir un point rapide sur la situation dans sa propre langue.

Concrètement, répondre à ce besoin implique plusieurs défis scientifiques : extraire un résumé d'une vidéo en se basant sur l'information vidéo (résumé de vidéo) et sonore (résumé de texte), et traduire le flux sonore dans une autre langue (traduction parole-texte, ou même parole-parole). Un autre défi est la recherche du flux de traitement le plus performant, sachant que chaque étape entraîne des erreurs : comment tenir compte des critères de résumé vidéo (rendre compte des différents lieux filmés, et peut-être ne pas insister trop longtemps sur le présentateur du journal) et des critères de résumé textuel

Volume 4 : Démonstrations et résumés d'articles internationaux, pages 53–56.
HAL : hal-02768513.

(tenir compte des idées exprimées) et marier ces critères ? Doit-on transcrire et traduire le résumé, ou bien transcrire, traduire la totalité de la vidéo, puis la résumer, etc. ? Toute la combinatoire est possible.

Le projet AMIS [1] a cherché à répondre à ces questions. Au-delà, un objectif était aussi de permettre à l'utilisateur de comprendre les différents points de vue via la comparaison en termes de sentiments et d'opinions de plusieurs vidéos comparables ; mais la démonstration ne concernera que les aspects de traduction et résumé. Ce projet de type Chist-Era implique plusieurs partenaires internationaux experts en extraction d'information vidéo (AGH [2], Cracovie, Pologne), en résumé de texte (LIA [3], Avignon, France), en transcription et traduction automatique de parole (Loria [4], Nancy, France), en test utilisateur (Deusto [5], Bilbao, Espagne). Nous décrivons ci-après les différentes configurations de résumé-vidéo-texte-traduction (Section 2) issues du projet, puis nous présentons la démonstration que nous montrerons pendant la conférence (Section 3).

2 Le processus de résumé/traduction

Le projet a abouti à quatre scénarios de fabrication d'un résumé vidéo dans une autre langue. Les scénarios reposent sur l'utilisation de la reconnaissance de la parole, de la traduction de texte, et sur diverses approches de résumé automatique : à partir des images, de l'audio ou du texte. Ces différents composants ont d'abord été évalués individuellement sur la base des métriques d'évaluation usuelles (Smaïli *et al.*, 2019). Puis les scénarios ont été comparés sur la base de retours utilisateur concernant la compréhensibilité du résumé, et son adéquation par rapport à la vidéo source.

Les quatre scénarios mis en œuvre et schématisés sur la figure 1 sont les suivants :

— Sc1 repose sur un résumé à partir des images vidéo. L'audio des segments sélectionnés est ensuite traitée par la reconnaissance de la parole, et le résultat est traduit dans la langue cible.
— Sc2 repose sur un résumé à partir des informations audio. Les segments obtenus sont ensuite traités par la reconnaissance de la parole et traduits dans la langue cible.
— Sc3 repose sur la reconnaissance de parole de l'émission complète et sa traduction dans la langue cible. Le résumé est appliqué sur le texte traduit. Enfin les segments vidéo correspondant aux phrases ou portions de phrase du résumé sont extraits et concaténés pour fournir le résumé de l'émission.
— Sc4 est similaire à Sc3, mais le résumé est réalisé sur le résultat de la reconnaissance de parole, puis traduit dans la langue cible.

3 La démonstration

Au cours de la démonstration, nous montrerons des vidéos en langue arabe, et leur résumé en langue anglaise (voir Figure 2). Du fait des temps de calcul non temps réel, le processus de résumé/traduction

1. Access Multilingual Information opinionS, `http://deustotechlife.deusto.es/amis/`
2. Akademia Górniczo-Hutnicza, `http://deustotechlife.deusto.es/amis/partner/3`
3. Laboratoire d'Informatique d'Avignon, `http://deustotechlife.deusto.es/amis/partner/5`
4. Laboratoire Lorrain en Informatique et ses Applications, `http://deustotechlife.deusto.es/amis/partner/2`
5. Université de Deusto, `http://deustotechlife.deusto.es/amis/partner/4`

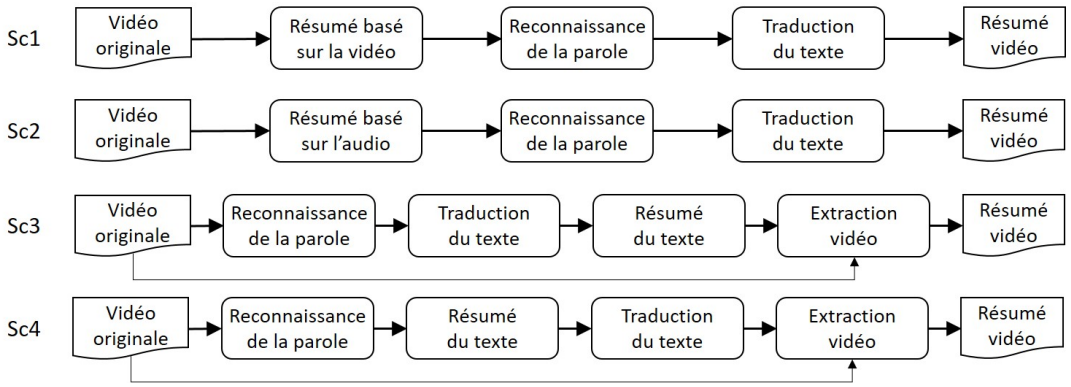

FIGURE 1 – Architectures pour la fabrication de résumés de vidéos dans une autre langue.

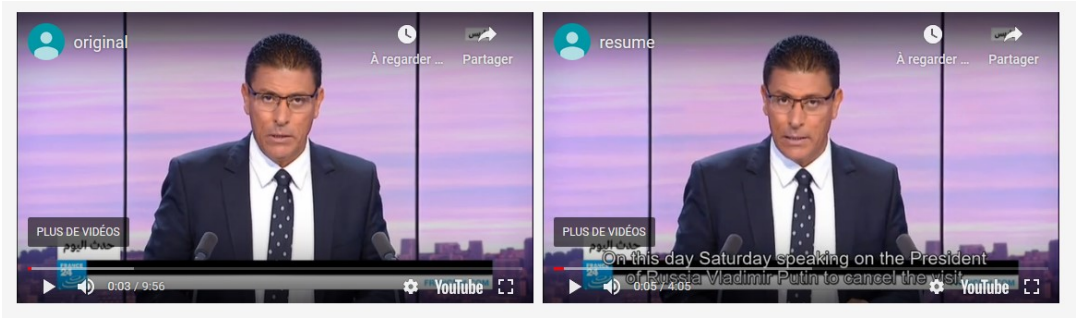

FIGURE 2 – Capture d'écran de la démonstration, à gauche la vidéo initiale en arabe (10 minutes), à droite la vidéo résumée (4 minutes) et les sous-titres en anglais.

ne se fera pas en direct. La traduction sera donnée sous forme de sous-titres. Le contenu arabophone des vidéos sera décrit. La démonstration sera l'occasion de relever les passages qui ont été bien sélectionnés et traduits, et ceux pour lesquels les résultats sont à améliorer. Les différentes vidéos montreront des résultats suite aux traitements des quatre scénarios. La démonstration sera aussi l'occasion de revenir sur les résultats publiés (Menacer *et al.*, 2019; Smaïli *et al.*, 2018), et de les commenter sur des exemples, et enfin de répondre à toute question sur les aspects scientifiques et techniques des méthodes utilisées pour la transcription et la traduction.

Remerciements

Nous remercions Chist-Era pour avoir financé ce travail via le projet AMIS (Access Multilingual Information opinionS). Cette démonstration a été organisée par le LORIA, mais se base sur les travaux de l'ensemble des partenaires AGH, DEUSTO, LIA, et LORIA.

Références

MENACER M. A., GONZÁLEZ-GALLARDO C. E., ABIDI K., FOHR D., JOUVET D., LANGLOIS D., MELLA O., SADAT F., TORRES-MORENO J. M. & SMAÏLI K. (2019). Extractive Text-Based Summarization of Arabic videos : Issues, Approaches and Evaluations. In *ICALP : International*

Conference on Arabic Language Processing, volume Communications in Computer and Information Science book series (CCIS, volume 1108), p. 65–78, Nancy, France : Springer.

SMAÏLI K., FOHR D., GONZÁLEZ-GALLARDO C., GREGA M., JANOWSKI L., JOUVET D., KOMOROWSKI A., KOZBIAL A., LANGLOIS D., LESZCZUK M., MELLA O., MENACER M. A., MENDEZ A., LINHARES PONTES E., SANJUAN E., SWIST D., TORRES-MORENO J.-M. & GARCIA-ZAPIRAIN B. (2018). A First Summarization System of a Video in a Target Language. In *MISSI 2018 - 11th edition of the International Conference on Multimedia and Network Information Systems*, p. 1–12, Wroclaw, Poland.

SMAÏLI K., FOHR D., GONZÁLEZ-GALLARDO C.-E., GREGA M. L., JANOWSKI L., JOUVET D., KOŹBIAL A., LANGLOIS D., LESZCZUK M., MELLA O., MENACER M.-A., MENDEZ A., PONTES E. L. L., SANJUAN E., TORRES-MORENO J.-M. & GARCIA-ZAPIRAIN B. (2019). Summarizing videos into a target language : Methodology, architectures and evaluation. *Journal of Intelligent and Fuzzy Systems*, **1**, 1–12.