

Extraction et gestion d'informations pour la construction d'une base vidéo d'apprentissage

Alain Simac-Lejeune*

*Litii

alain.simaclejeune@litii.com,
<http://www.litii.com>

Résumé. Indexer une vidéo consiste à rattacher un ou plusieurs concepts à des segments de cette vidéo, un concept étant défini comme une représentation intellectuelle d'une idée abstraite. L'indexation automatique se base sur l'extraction automatique de caractéristiques fournies par un système de traitement d'images. Cependant, il est nécessaire de définir les index ou concepts. Pour cela il faut définir le lien qui existe entre ces caractéristiques et ces concepts. Ce qui sépare les caractéristiques extraites sur lesquelles se base l'indexation automatique et les concepts est appelé fossé sémantique qui est le manque de concordance entre les informations que les machines peuvent extraire depuis les documents numériques et les interprétations que les humaines en font. La définition d'un concept peut être faite automatiquement si l'on dispose d'une base d'apprentissage liée au concept. Dans ce cas, il est possible "d'apprendre" le concept de manière statistique. Mais la construction de cette base d'apprentissage nécessite de faire intervenir un utilisateur ou un expert applicatif. En fait, il s'agit de s'appuyer sur ses connaissances pour extraire des segments vidéo représentatifs du concept que l'on souhaite définir. On peut lui demander d'indexer manuellement la base d'apprentissage, mais cette opération est longue et fastidieuse. Dans cet article, nous proposons une méthode qui permet d'extraire l'expertise pour que l'implication de l'expert soit la plus simple et la plus limitée possible.

1 Introduction

Actuellement, de nombreuses recherches traitent de l'analyse et de la reconnaissance d'activité. Ces recherches portent sur le développement de traitements automatiques de vidéo en terme d'extraction d'informations, de modélisation de ces informations et de reconnaissance automatique. La construction et l'évaluation de ces traitements requièrent des corpus vidéo représentant de nombreuses heures. Ces vidéos sont généralement annotées manuellement. Dans le cas de vidéo de longue durée, l'annotation manuelle est source d'erreurs, non reproductible et chronophage. De plus, la qualité des annotations dépend grandement de l'expertise de l'utilisateur. L'association de cette expertise à des traitements automatiques facilite cette tâche et représente un gain de temps.

La difficulté d'intégration de l'utilisateur dans la boucle de définition ne réside pas directement