

---

# Modélisation des (méta)données hétérogènes et filtrage des contenus de vidéosurveillance : application au Forensic

**Franck Jeveme Panta**

*Université de Toulouse III, Paul Sabatier  
Institut de Recherche en Informatique de Toulouse (IRIT), CNRS  
118 Route de Narbonne  
31062 TOULOUSE CEDEX 9*

*Franck.Panta@irit.fr*

---

*MOTS-CLÉS : Vidéosurveillance, métadonnées, filtrage, hétérogénéité, forensic.*

*KEYWORDS: CCTV system, metadata, filtering, heterogeneity, forensic.*

*ENCADREMENT. Florence Sèdes (PR) et André Péninou (MCF)*

---

## 1. Contexte

Au cours des dernières années, le taux de criminalité a augmenté très rapidement dans le monde. Le nombre de caméras de vidéosurveillance censées assurer la sécurité des citoyens dans les grandes villes croît régulièrement et de manière significative. Ces caméras génèrent une masse importante de données vidéos à analyser afin d'élucider les différents crimes (terrorisme, homicide, enlèvement, ...) survenus. Le processus actuellement utilisé par les enquêteurs (en France), consiste à appliquer des traitements vidéo (détection de visages, détection de véhicules, détection et lecture des plaques d'immatriculation...) sur l'ensemble des séquences vidéos récupérées en amont et préalablement indexées (conversion au bon format, géolocalisation des caméras, gestion des horodatages...). Les traitements sont appliqués sur l'ensemble des vidéos ou uniquement sur un sous-ensemble géographique ou temporel. Après ces traitements et une extraction des données, les enquêteurs effectuent des recherches (véhicules, plaques, personnes, visages...), visionnent les séquences d'intérêt et lancent de nouveaux traite-

ments. Les affaires récentes (terrorisme, enlèvement, homicide) ont nécessité l'analyse de plusieurs dizaines de milliers d'heures de vidéo. Le gain de temps fourni par les outils d'analyse actuels reste insuffisant dans un contexte opérationnel. Beaucoup des vidéos analysées pour une enquête s'avèrent inexploitable (par ex., problème de luminosité) pour l'enquête. Écarter ces séquences inéligibles permettra d'optimiser le temps d'exploitation. Cette thèse a pour objectif de proposer une (méta)modélisation et les mécanismes afférents de filtrage de grandes collections de vidéo liées à la recherche de preuves à posteriori. Cette contribution basée sur les (méta)données passe par la révision de la norme ISO 22311 qui décrit les exigences fonctionnelles à satisfaire par les systèmes de vidéosurveillance.

## 2. État de l'art

L'analyse vidéo automatique est actuellement confrontée à trois défis majeurs : (i) problème de gestion du grand nombre de caméras largement distribuées, (ii) difficulté de stockage et d'organisation de l'énorme quantité de données générées par ces caméras, (iii) analyse, annotation et recherche demeurent très coûteuses en temps.

De nombreux travaux ont été menés dans le but de proposer des solutions aux problèmes d'analyse vidéo ; notre objectif n'est pas d'en faire un exposé exhaustif ni d'apporter une contribution à l'analyse de contenu proprement dite : le verrou adressé est celui qui consiste à enrichir la modélisation des (méta)données, actuellement pas ou peu exploitées dans ce domaine.

Une méthode efficace et efficiente pour la construction et l'extraction des métadonnées descriptives des vidéos Web est présentée dans (Algur SP, et al., 2014) qui propose un modèle de métadonnées descriptives pouvant faciliter le traitement des contenus vidéos. Ce modèle est orienté classification des vidéos et nécessite d'autres spécifications pour être utilisé dans notre contexte. Un framework utilisant des informations géographiques pour la récupération des vidéos sur le web est présenté dans (Han Z. et al., 2016). Les auteurs décrivent et utilisent des données géographiques liées à la vidéo, telles que la localisation vidéo, le champ de vue de la caméra et les trajectoires pour effectuer des recherches dans les vidéos. Cependant, la qualité d'images/vidéos n'est pas prise en compte.

Les récents travaux de notre équipe se sont focalisés sur la modélisation des métadonnées pour la gestion distribuée des documents multimédias et l'interopérabilité des systèmes de vidéo surveillance. (Manzat, 2013) propose un format de métadonnées associées à tout type de contenu multimédia et permettant à plusieurs systèmes d'échanger des données et de les utiliser sans effectuer de manipulations spéciales. Ce format de métadonnées a été validé dans le projet LINDO<sup>1</sup>. Une étude des métadonnées utiles pour la recherche dans les collections de vidéos est présentée dans (Codreanu, 2015). Cette étude a été appliquée dans le cadre du projet ANR METHODEO<sup>2</sup>. L'approche

---

1. <http://lindo-itea.eu/>

2. <http://www.agence-nationale-recherche.fr/Projet-ANR-10-SECU-0006>

utilisée dans (Codreanu et al., 2015) vise à aider les opérateurs humains de vidéosurveillance dans la recherche de séquences vidéo d'intérêt en leur proposant un ensemble de caméras susceptibles d'avoir filmé une scène recherchée. Une extension de cette approche est décrite dans (Panta et al,2016).

Globalement, les solutions proposées dans la littérature utilisent des approches d'interrogation de contenus basées sur des métadonnées (techniques, liées aux caméras, spatio- temporelles) mais ne proposent pas d'interrogation basée sur le contenu via les descripteurs de qualité images/vidéos ou d'autres caractéristiques liées aux contenus.

### **3. Problématique**

Un rapide état de l'art montre donc la nécessité d'exploiter les informations concises des contenus vidéo à partir de tous les éléments disponibles, a priori ou a posteriori en fonction des (méta)données potentiellement élicitable. Ces informations constituent des métadonnées issues d'algorithmes d'analyse de contenu (caractéristiques de bas niveau, présence et/ou nombre de personnes dans la scène, ...) et peuvent être modélisées selon différents niveaux de sémantique et de granularité. La collaboration des différents niveaux de métadonnées est un verrou essentiel dans notre approche. Cette approche consiste à effectuer un filtrage de contenus vidéo avec le meilleur compromis qualité/précision/temps de réponse, en combinant les métadonnées techniques, les métadonnées décrivant le mouvement et le champ de vue de la caméra, les métadonnées issues d'algorithmes d'analyse de contenu, les métadonnées de contexte (réseaux sociaux, open data, géolocalisation indoor/outdoor, jour/nuit). Une telle approche s'apparente à un problème de modélisation multicouches et est caractérisée par l'hétérogénéité des données dont elle est issue.

Nous souhaitons lever les verrous relatifs à (i) la proposition de "descripteurs" de qualité, (ii) l'interopérabilité des métadonnées associées aux contenus vidéo, (iii) la recherche dans de grands volumes de données vidéo sans recourir aux contenus intégraux, ces problématiques visant à mettre en oeuvre le filtrage de contenus pertinents pour les enquêteurs.

### **4. Actions réalisées**

Pour lever les verrous mentionnés dans la problématique, notre méthodologie consiste dans un premier temps à faire une étude des métadonnées de contenus utiles pour le filtrage négatif/positif des contenus vidéo. Cette tâche consiste à extraire les caractéristiques des contenus vidéo et les modéliser selon les différents niveaux de granularité (frame, segment, vidéo) et de sémantique.

Notre travail s'applique dans le contexte de la recherche des preuves à posteriori (forensic) et se déroule dans le cadre du projet ANR FILTER2<sup>3</sup>. Les caractéristiques sont

---

3. FILtrage négaTif des contEnus de vidéoprotection

extraites des vidéos en fonction des différents traitements effectués dans les enquêtes policières. Trois traitements sont retenus dans le cadre de FILTER2 : la détection de visages, la détection et la lecture automatique des plaques, et la détection des véhicules. Nous avons donc défini les différents niveaux de métadonnées de contenus suivants : (i) les métadonnées liées aux pixels (Low level) : éclairage, luminosité, contraste, taux de bruit, taux de mouvement, etc. ; (ii) les métadonnées liées aux objets (Medium level) : présence des personnes et véhicules, nombre de personnes, etc. ; (iii) les métadonnées liées à la description de la scène (High level) : intempéries, nuit/jour ,etc. Nous proposons actuellement un modèle de données intégrant toutes ces métadonnées de contenus et celles mentionnées dans l'état de l'art.

## 5. Actions futures

Dans le but de mettre en place le filtrage négatif/positif des contenus, les prochaines directives de travail sont : (i) la définition des critères de qualité et d'utilisabilité des vidéos grâce à la combinaison des différentes métadonnées, (ii) la généralisation des approches de modélisation des métadonnées, (iii) la mise en oeuvre des algorithmes de requête basés métadonnées pour le filtrage des contenus.

## Bibliographie

- Algur SP, Bhat P, Jain S. Metadata Construction Model for Web Videos : A Domain Specific Approach. *International Journal of Engineering and Computer Science*. 2014 Dec 28;3(12).
- Manzat, Ana-Maria. "Contribution à la modélisation des métadonnées associées aux documents multimédias et à leur enrichissement par l'usage." PhD diss., École Doctorale Mathématiques, Informatique et Télécommunications (Toulouse); 142547247, 2013.
- Codreanu, Dana. "Modélisation des métadonnées spatio-temporelles associées aux contenus vidéos et interrogation de ces métadonnées à partir des trajectoires hybrides : application dans le contexte de la vidéosurveillance." PhD diss., Université de Toulouse, Université Toulouse III-Paul Sabatier, 2015.
- Codreanu D, Peninou A, Sedes F. Video Spatio-Temporal Filtering Based on Cameras and Target Objects Trajectories–Videosurveillance Forensic Framework. In *Availability, Reliability and Security (ARES), 2015 10th International Conference on 2015 Aug 24* (pp. 611-617). IEEE.
- Panta FJ, Sèdes F. Mobile objects in indoor environment : Trajectories reconstruction. In *Proceedings of the 14th International Conference on Advances in Mobile Computing and Multi Media 2016 Nov 28* (pp. 332-336). ACM.
- Han Z, Cui C, Kong Y, Qin F, Fu P. Video data model and retrieval service framework using geographic information. *Transactions in GIS*. 2016 Oct 1;20(5) :701-17.