L'évaluation d'algorithmes d'analyse vidéo – Quelques pistes

Jean-François GOUDOU[1], Louise NAUD[1], Laurent GIULIERI[2,] Jaonary RABARISOA[3], Olivier PIETQUIN[4], Dana CODREANU[5], Dijana PETROVSKA[6]

[1]THALES Theresis, 1 avenue Augustin Fresnel, 91767 Palaiseau Cedex

[2]Digital Barriers, E.Golf-Park, bâtiment A, 950 avenue Roumanille, BP 80187, 06904 Sophia-Antipolis Cedex

[3]CEA SACLAY - Nano-INNOV, Bât. 861 - Point courrier 173, 91191 Gif-sur-Yvette Cedex

[4]Supélec Campus de Metz, 2 rue Edouard Belin, 57070 Metz

[5]IRIT, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse

[6]Telecom SudParis, 9 rue Charles Fourier, 91011 Evry

jean-francois.goudou@thalesgroup.com, louise.naud@thalesgroup.com, laurent.giulieri@digitalbarriers.com, jaonary.rabarisoa@cea.fr, olivier.pietquin@supelec.fr, dana.codreanu@irit.fr, dijana.petrovska@telecom-sudparis.edu

Résumé – Le projet METHODEO vise à définir une méthodologie générique d'évaluation des algorithmes d'exploitation des enregistrements dans le cadre des enquêtes judiciaires. L'idée principale est d'aider les opérateurs à prévoir avec une bonne précision le comportement des algorithmes qui accomplissent une certaine fonctionnalité sur des données réelles (dans un environnement réel) en faisant une comparaison entre la description des nouvelles données avec les descriptions des classes de vidéos sur lesquelles les algorithmes ont déjà été évalués.

Abstract – The METHODEO project aims at the definition of a new generic evaluation methodology for the algorithms dedicated to the exploitation of recorded videos for forensic issues. The main focus is on helping the operators to forecast with a good precision the behavior of video analytics on real data, in a real environment, by comparing features extracted from the new video with features describing several classes of recorded videos with asserted video analytics performances.

1. Objectifs du projet

Alors que le nombre de déploiements de systèmes de vidéo-protection augmente régulièrement et qu'ils sont constitués par de plus en plus de cameras, les opérateurs de vidéo protection ont des difficultés grandissantes d'exploitation dues aux grandes quantités de données vidéos stockées chaque jour. C'est particulièrement le cas lorsque, dans le cadre d'une enquête judiciaire, le temps passé par les opérateurs pour retrouver des données suspectes est extrêmement long alors que les délais d'obtention sont critiques.

Les activités de recherche se sont multipliées afin de répondre à ces besoins. Les fruits de ces recherches arrivent à maturité et des algorithmes et des produits d'analyse vidéo émergent sur le marché de la sécurité. A la différence des systèmes logiciels classiques il est difficile de connaitre les performances exactes de ces nouveaux outils à cause de la très grande variabilité des vidéos d'entrée en termes d'espace des paramètres. Il est pourtant capital de pouvoir prévoir avec une bonne précision leur comportement sur des données réelles. C'est le rôle des systèmes d'évaluation qui doivent permettre d'indiquer les performances d'un algorithme pour une **fonctionnalité** attendue par un opérateur dans un **environnement** donné. Les systèmes d'évaluation existants caractérisent une scène à partir d'une dizaine de paramètres (luminosité, présence d'ombres, etc.) ce qui ne permet pas de couvrir l'ensemble des situations et donc finalement de décrire complètement les performances d'un algorithme. De plus les opérateurs de vidéo-protection ne disposent actuellement d'aucune méthodologie pour évaluer les outils proposés par les différents fournisseurs. Evaluation rendue difficile car les métadonnées de sortie qui décrivent les résultats des outils d'analyse vidéo adoptées par les fournisseurs ne sont pas toujours compatibles. Cette limitation s'intègre de manière plus globale à la problématique d'interopérabilité des systèmes de vidéo-protection.

En premier lieu il s'agit de comprendre précisément les besoins des opérateurs durant une enquête et de les traduire en termes de **fonctionnalité** afin d'identifier les algorithmes qui y répondent. La caractérisation de l'**environnement** est ensuite considérée. L'objectif est de structurer une base de données vidéo d'évaluation via la description la plus précise, exhaustive et objective possible

des vidéos qui la compose en s'appuyant d'une part sur les annotations des vidéos (boites englobantes, descriptions textuelles, etc.) qui permettent l'évaluation des performances des algorithmes et d'autre part sur des données de caractérisation/indexation du contenu image (mesure de contraste, couleur, etc.). Un apprentissage sur ces 2 types de données est alors envisagé afin d'extraire des clusters de vidéo dont l'environnement/contexte de la scène est similaire. Enfin un intérêt particulier est porté sur la standardisation des métadonnées générées par les outils d'analyse vidéo. Celles-ci suivront le dictionnaire définit par la norme ISO 22311 dont l'objectif est d'assurer l'interopérabilité des systèmes de vidéo protection pour faciliter le travail d'enquête des forces de l'ordre et de la justice.

Un état de l'art constitué des campagnes d'évaluation en analyse vidéo TRECVID / ETISEO est proposé en première partie de cet article. La seconde partie décrit la méthodologie proposée où l'aspect standardisation est introduit (norme 22311). Enfin la caractérisation de l'environnement/contexte de la scène par apprentissage sur les données de description est abordée en dernière partie.

2. L'état de l'art en évaluation

L'évaluation est une partie importante de tout processus scientifique. Comme souligné dans [5, chap.1], chercheurs, industriels et utilisateurs se posent constamment les questions suivantes:

- Comment mesurer le progrès de la recherche?
- Comment savoir si un algorithme a des meilleures performances qu'un autre ?
- Est-ce que les données sur lesquels on évalue sont représentatifs et de taille suffisante ?
- Comment vont se généraliser les résultats sur d'autres données ?
- Est-ce que les bons résultats sont reproductibles par tous?
- Pour quelles applications ces algorithmes sont utiles ?

Les réponses à ces questions s'obtiennent au travers des évaluations. Faut-il encore que ces évaluations suivent des règles bien définies ! Il est notamment important d'utiliser des bases de données publiquement disponibles, afin que ces systèmes soient aussi évalués sur des bases séquestres et avec des protocoles publics et diffusés dans la communauté. Lorsque les technologies impliquées relèvent des problématiques de reconnaissance de forme - il faut tenir compte des bases de données qui sont nécessaire pour développer et tester les systèmes (de reconnaissance, classification, indexation,..)

Lorsqu'il s'agit de traiter des vidéos, l'évaluation peut devenir relativement complexe en fonction de la tâche que l'on considère et de la complexité des donnés à traiter. Dans le projet METHODEO on est principalement intéressé dans le traitement des séquences vidéo (provenant par exemple des caméras de vidéosurveillance). Pour la majorité des tâches, une chaîne de traitement composé de plusieurs algorithmes est nécessaire. Ainsi si l'on veut évaluer l'importance de chaque maillon de la chaine, il faut faire des évaluations spécifiques pour chaque partie. Par exemple si on veut reconnaitre une personne par son visage, on doit d'abord procéder à une étape de détection de visage, et éventuellement de détection de points caractéristique avant d'arriver à l'étape de comparaison de deux visages.

Il est important de souligner que la partie préparation de corpus représente une partie qui est très couteuse, et ceci tout aussi bien la partie enregistrement des données et annotation (vérité terrain). La partie annotation peut être faite soit par un expert humain, sot de manière semi-automatique. Il est aussi crucial de bien définir les conditions d'enregistrements des corpus, car ce sont ces spécifications qui vont déterminer de l'utilité et de l'adéquation des corpus pour des fonctionnalités bien spécifiques.

Les évaluations actuelles se basent sur des corpus suivants :

- Corpus orientés vidéosurveillance (tels que ETISEO, TrecVid): Ils regroupent les bases qui présentent des scènes complexes, avec une diversité d'objets à détecter (y compris des voitures et des personnes). Les personnes sont plutôt de taille réduite.
- Corpus (images fixes principalement) orientés détection et classification d'objets.
- Corpus orientés biométrie, avec différentes caractéristiques (modalités) biométriques. Les modalités biométriques qui sont les plus utiles pour l'analyse des vidéos a posteriori sont la silhouette et le visage.
- Corpus vidéosurveillance /biométrie : qui combinent ces deux fonctionnalités sont plus rares. On peut citer comme exemple le corpus MBGC, qui dans la partie « Vidéo Challenge » fournit les données qui peuvent être exploitables pour combiner vidéosurveillance et biométrie conjointement.

La difficulté de l'évaluation dans le cas de recherche de preuve a posteriori provient du fait que l'on a besoin de connaitre des performances d'un algorithme pour une **fonctionnalité** attendue par un opérateur dans un **environnement** donné. Ainsi le projet METHODEO propose d'établir une nouvelle méthodologie d'évaluation des algorithmes d'exploitation des enregistrements dans le cadre des enquêtes judiciaires.

3. La méthodologie METHODEO

3.1 Principe de la méthodologie

Scénario

Recherche judiciaire sur un ensemble de vidéos. Avec quels algorithmes / paramètres ?

Préparation

Il existe une base de données générique, caractérisée grâce à notre méthodologie, sur laquelle les autorités ont évalué tous les algorithmes dont elles disposent.

Des clusters de vidéos similaires sont constitués, la similarité étant calculée à partir du contexte, des descripteurs et des annotations.

Chaque couple cluster – fonctionnalité se voit attribuer une chaine de traitement optimale grâce à notre méthodologie.

Pendant la requête

Les autorités vont réquisitionner des vidéos. Les vidéos les plus similaires dans la BDD sont identifiées grâce à une mesure de similarité.

Les algorithmes les plus performants sur la BDD seront sélectionnés pour une fonctionnalité et un contexte donnés.

3.2 Indexation suivant le contenu

L'objectif de l'étape d'indexation par le contenu est de définir des descripteurs susceptibles de discriminer des vidéos entre elles en fonction de la réponse d'un algorithme donné. Il existe à ce dessein deux catégories de descripteurs :

- Les descripteurs bas niveau, basés sur le contenu brut de l'image,
- Les descripteurs haut niveau, orientés vers la sémantique du contenu de la scène.

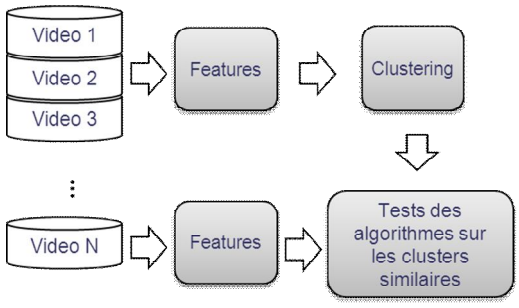

Figure 1 : Méthode de comparaison d'algorithmes par le contenu vidéo

exploitables pour associer des vidéos selon leurs réponses algorithmiques. En effet, une même scène filmée à la lumière du jour puis à la lumière d'un éclairage public la nuit ne demandera pas nécessairement le même algorithme pour détecter une intrusion. Corrélativement, une scène filmée simultanément par deux caméras distinctes peut requérir deux algorithmes différents pour suivre les objets mobiles qui y sont présents.

L'approche retenue ici pour le projet METHODEO consiste à générer des descripteurs bas niveau, basés sur la couleur (MPEG7), des points d'intérêts (SIFT, SURF, FAST), ou encore de texture (matrice de cooccurrence). Grâce à des métriques adaptées, une classification est alors effectuée sur l'espace mixte des descripteurs ; les vidéos appartenant au même groupe à l'issue de cette étape répondent de manière similaire à un même algorithme.

Parmi les descripteurs nous étudions ceux qui sont liés aux objets présents dans la scène et ceux qui sont liés au contexte. Les premiers vont donner une description objective, indépendant des algorithmes à évaluer et des objets eux-mêmes. Par exemple, on décrit le mouvement des objets en utilisant les informations comme le flux optique, la densité de présence dans l'image en regardant le taux de pixels occupés, la texture en quantifiant les descripteurs des points d'intérêt, etc. Ces autres types de descripteurs servent à décrire le contexte. On retrouve ici tous les types de descripteur image bas niveau précédemment cités.

Lorsque l'on cherche à savoir sur une nouvelle vidéo quel algorithme répondrait le mieux, on peut alors calculer ses descripteurs et en déduire de quelle classe de vidéos elle se rapproche le plus. Par voie de fait, on peut alors déduire l'algorithme qui répondra le mieux sur cette vidéo.

3.3 Annotations et standardisation

Les métadonnées vidéo sont des données de description de contenu vidéo. Ces données peuvent être aussi bien des descriptions textuelles, boites englobantes, etc. que des données extraites automatiquement des images représentées sous forme de descripteurs (mesure de contraste, colorimétrie, etc.) tels que définis dans la section 3.2. Ces métadonnées sont renseignées par les outils/algorithmes d'analyse vidéo. Elles seront par contre exploitées par les outils de recherche ou d'indexation vidéo. Evaluer des algorithmes revient donc à comparer leurs sorties, c'est-à-dire les métadonnées qu'ils ont générés, à des vérités terrain, c'est-à-dire des métadonnées références préalablement annotées par l'homme. La comparaison des performances d'algorithmes demande qu'un dictionnaire commun soit défini afin d'assurer des sorties et vérités terrain communes. Des standards existent (ONVIF, PSIA) mais ne sont pas compatibles entre eux. La norme ISO 22311 a pour objectif de définir un standard international afin d'assurer l'interopérabilité des systèmes de vidéo protection pour faciliter le travail d'enquête des forces de l'ordre et de la justice. Un dictionnaire de métadonnées existe, il permet la description d'un système de vidéo protection et des évènements qui peuvent se produire. L'interopérabilité des systèmes de vidéo protection inclut l'interopérabilité des

outils de recherche et d'analyse vidéo, aussi nous nous proposons d'utiliser le dictionnaire fourni par la norme et de l'enrichir.

3.3.1 Les métadonnées / norme 22311

La Norme ISO 22311 vise à faciliter l'interopérabilité des systèmes de vidéo protection en définissant un format générique d'export des données. La norme propose d'archiver les données de manière hiérarchique dans des fichiers, dossiers et groupes de dossiers. Le rangement en fichiers et dossiers est fait conformément à des créneaux temporels de répertoire (DTS) et créneaux temporels de fichiers (FTS) (le temps est donné en temps universel coordonné (UTC)). Chaque fichier contient des données (audio/vidéo) provenant de plusieurs sources (cameras), un index par contenu (audio/vidéo) permettant l'accès précis à toute image et toute heure spécifiques et des métadonnées pour chaque source (caméra) et dossier. La norme n'a pas comme objectif de définir des nouveaux modèles de métadonnées mais plutôt de proposer une structure de format d'encapsulation qui se base sur des standards déjà existants, notamment MPEG, JPEG, CEI/TC et SMPTE. Les éléments qui doivent obligatoirement être fournis par tous les systèmes de vidéosurveillance afin d'assurer un minimum d'interopérabilité (visualisation des vidéos) sont : nom et profil du codec, nom du conteneur, résolution vidéo, nombre d'images vidéo (en images/secondes), heure et date de l'enregistrement et heure et date de la caméra.

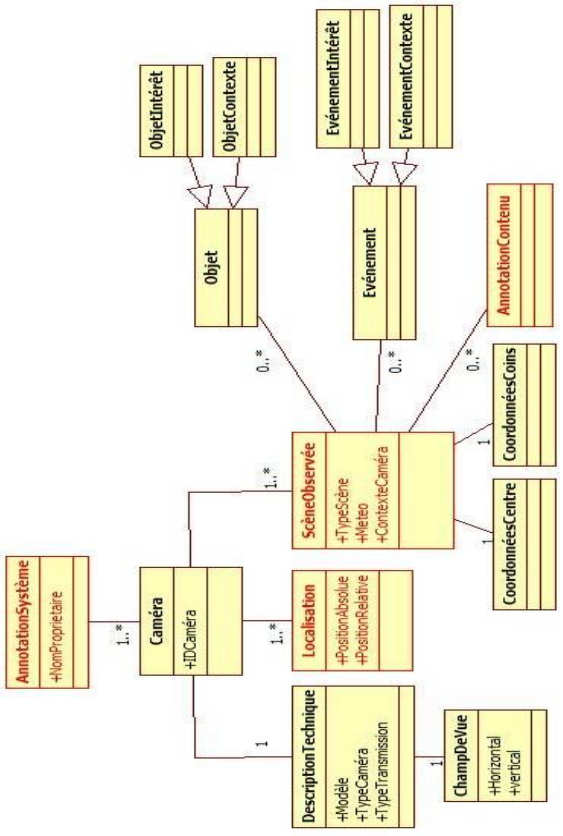

Figure 2 : Format d'annotation

Ensuite la norme propose des dictionnaires de métadonnées concernant les capteurs et les événements. Les métadonnées concernant les capteurs contiennent des éléments décrivant : des informations génériques sur le dispositif (modèle, fabricant etc.), des informations optiques (distance focale, champ de vue etc.), la localisation du dispositif et de la scène observée (coordonnées géographiques, etc.), des informations temporelles, etc. Les événements sont décrits en termes d'heure de début, localisation et liens vers des dictionnaires d'événements, vers d'autres événements et vers les capteurs qui contiennent des informations sur l'événement.

3.3.2 Définition du dictionnaire de données

Le format proposé dans METHODEO utilise des éléments du dictionnaire de la norme 22311 (e.g. coordonnées spatiales et temporelles). Le dictionnaire de métadonnées est complété avec les éléments pertinents (au sens de l'analyse vidéo) de description des collections vidéo et de description du réseau de caméras. Plusieurs éléments décrivant les collections vidéo doivent être pris en compte dans la définition de la nouvelle méthodologie. Un schéma simplifié de notre format est proposé Figure 2 ou les éléments originaux par rapport à la norme sont représentés en rouge. Le dictionnaire doit permettre de décrire les éléments techniques du système de vidéo protection, les éléments bas niveau et haut niveau (sémantique) du contenu vidéo (voir Section 3.2) et les éléments de haut niveau (sémantiques) de la scène. Le but de cette étude est de recenser de la façon la plus complète possible ces éléments.

L'annotation du système est associée à plusieurs annotations caméras. Pour chaque caméra (capteur dans la norme) on renseigne des informations techniques (e.g. le modèle, le codec, la qualité de l'enregistrement), la localisation (géographique et par rapport au réseau de caméras) et on associe une scène observée, en décrivant les coordonnées de la scène (qui peuvent être dynamiques), le type de scène, le contexte de la scène etc. Aux vidéos correspondant à chaque scène on va associer les annotations des objets et des événements et les descripteurs de bas niveau.

Les annotations vidéo peuvent être classifiées en utilisant plusieurs critères. Les plus pertinents retenus dans le contexte de METHODEO sont:

Le type de description

- *Annotations du système de vidéo surveillance* (e.g. localisation absolue (géographique) et relative (par rapport au système) des caméras, description technique des caméras)
- *Annotations de la scène* (e.g. coordonnées de la scène observée par chaque caméra, conditions météo, type de scène)

- *Annotations du scénario* (e.g. description des objets et des événements)

Le stockage et la synchronisation

- *Annotations associées au système* (une description par système)
- *Annotations associées à la caméra* (une description par caméra)
- *Annotations associées au flux vidéo/audio* (description du contenu vidéo/audio par image ou par image clé)

Leur mutabilité

- *Annotations statiques* (e.g. caractéristiques des caméras)
- *Annotations dynamiques* (e.g . qualité d'image, localisation des caméras mobiles)

La façon de les générer

- *Annotations manuelles* (e.g. description textuelle des objets)
- *Annotations semi automatiques* (e.g. algorithme de détection de personnes et intervention humaine pour corrections et enrichissement)
- *Annotations automatiques* (e.g. boites englobantes associées aux objets détectés)

La distinction entre la description de la scène et du scénario nous impose de séparer les entités objets en objets d'intérêt (e.g. personne, véhicule, groupe de personnes, groupe de véhicules) et objets de contexte (e.g. banques, arbres, objets sur les quais du métro) et entités évènement en événements d'intérêt (e.g. bagarre, vol de sac) et événements de contexte (e.g. ouverture d'une porte ou d'une barrière).

3.3.3 Les outils

Il existe un panel d'outils d'annotations, chacun ayant son propre format : VIPER-GT[1], LabelMe[2], Vatic[3], Anvil[4]. Ces outils ont été utilisés dans les différentes campagnes d'évaluation (TRECVid[5], Pets[6]) ou projets (CAVIAR[7], I-Lids[8], ETISEO[9]). Les formats existants décrivent **des objets** en utilisant des boites englobantes ou des polygones et en associant des attributs comme : des descriptions textuelles, dimension, position etc., **des événements** en associant des attributs comme : intervalle temporaire, objets participants, description textuelle etc. et des **métadonnées génériques** comme le format, l'encodage, la date de création etc.

[1] *http://viper-toolkit.sourceforge.net/*

[2] *http://labelme.csail.mit.edu/VideoLabelMe/*

[3] *http://mit.edu/vondrick/vatic/*

[4] *http://www.anvil-software.de/*

[5] *http://trecvid.nist.gov/*

[6] *http://www.cvg.rdg.ac.uk/slides/pets.html*

[7] *http://homepages.inf.ed.ac.uk/rbf/CAVIAR/*

[8] *http://www.homeoffice.gov.uk/science-research/hosdb/i-lids/*

[9] *http://www-sop.inria.fr/orion/ETISEO/*

Les processus d'annotation proposés ne sont pas encore adaptés aux besoins industriels. Des points bloquants sont toujours à considérer. Un premier concerne la gestion de la mobilité des caméras, des objets d'intérêt et des utilisateurs dans le cadre d'un réseau de caméras **Erreur ! Source du renvoi introuvable.**. Un second point bloquant est le temps d'annotation de la quantité de données vidéo nécessaire à une évaluation significative. Des équipes de recherche travaillent actuellement sur les aspects de semi-automatisation des processus d'annotation. Par exemple une solution consiste à utiliser des algorithmes de détection pour détecter au préalable les objets d'intérêt et ensuite corriger manuellement les objets mal annotés. Ceci n'est efficace que si les algorithmes de détection sont efficaces. Un autre moyen d'accélérer la génération des annotations est d'annoter uniquement les images clés. Les annotations manquantes sont ensuite interpolées ou extrapolées avec des méthodes d'interpolation avancée. Par exemple, certaine méthodes d'interpolation cherchent la meilleure trajectoire entre les images clés en utilisant l'apparence des objets annotés dans celles-ci. Une sélection optimale du nombre d'images nécessaires à l'annotation réduira également le temps d'intervention d'un annotateur humain. Les méthodes les plus utilisées sont celles basées sur l'apprentissage actif **Erreur ! Source du renvoi introuvable.**. L'idée est d'apprendre à partir d'annotations existantes les meilleures images à annoter pour un objet donné.

Faire une annotation collaborative peut aussi accélérer la génération des annotations. On peut par exemple segmenter une vidéo et demander à plusieurs personnes d'annoter les images clés de chaque segment en même temps. Une approche envisagée dans METHODEO est de faire une annotation hiérarchique. Les annotations haut niveau comme les évènements sont plus faciles à obtenir. On peut donc utiliser ces informations pour extraire des annotations bas niveaux comment les boites englobant des objets entre le début et la fin de l'évènement. On peut aussi exploiter les annotations sémantiques, comme la présence ou non d'un objet particulier, qu'on peut obtenir facilement dans avec les systèmes de gestion de vidéo actuel pour arriver à une segmentation au niveau pixelique des vidéos **Erreur ! Source du renvoi introuvable.**.

3.4 Apprentissage

La plupart des algorithmes utilisés dans le domaine de la vidéosurveillance sont destinés à exécuter des tâches de détections, reconnaissance ou de suivi dans des flux vidéo. Ces tâches sont très difficiles à spécifier autrement que par l'exemple. Ainsi, les algorithmes apprennent grâce à des données sur lesquelles un oracle humain a réalisé la tâche. Comme indiqué plus haut, une première étape pour la mise en œuvre et l'évaluation de tels algorithmes est donc la collection de bases de données annotées, l'annotation servant en partie à décrire la tâche que l'algorithme devra réaliser (détecter la présence ou pas d'une personne dans

la vidéo, détection de visage, reconnaissance de plaques minéralogiques etc.). Plus formellement, l'algorithme doit donc apprendre à associer des entrées (caractéristiques extraites des images ou des vidéos) à des labels (cadre autour d'un visage, numéro de plaque, position des personnes dans l'image). Tout l'enjeu réside dans la capacité de généralisation des algorithmes c'est-à-dire leur capacité à associer le bon label à une nouvelle entrée qui n'était pas présente dans la base de données ayant servi à l'apprentissage. Plusieurs problèmes sont alors à résoudre parmi lesquels trois principaux :

1) Comment choisir un algorithme qui va apprendre à réaliser la tâche grâce à un minimum de données (le coût de l'annotation rendant souvent problématique la collection de bases conséquentes)
2) Comment constituer correctement les bases de données pour qu'elles soient suffisamment représentative du problème et amène l'algorithme voir un panel assez large d'exemples.
3) Comment extraire des caractéristiques pertinentes des données pour qu'il soit possible à l'algorithme de discriminer les différents classes présentes. Plusieurs types de caractéristiques existent telles que celles mentionnées au paragraphe 3.2.

Bien qu'il existe une large littérature scientifique pour trouver des réponses théoriques à ces questions [1], il n'y a malheureusement pas de réponse universelle. Suivant la tâche à résoudre et les conditions d'acquisition des données, une réponse *ad hoc* doit être approchée. Ainsi, du fait de la nature intrinsèque du problème d'apprentissage, il est nécessaire d'adopter une démarche pragmatique pour trouver la meilleure combinaison algorithme-données-caractéristiques pour une tâche donnée. Une partie de la recherche réalisée dans le cadre du projet METHODEO a ainsi consisté à définir une telle méthodologie.

Le type d'apprentissage décrit jusqu'ici est dit « supervisé » du fait qu'il nécessite l'intervention d'un annotateur humain qui réalise la tâche. Dans la méthodologie proposée au paragraphe 3.2, une étape de regroupement automatique des données selon des métriques de similarité est aussi nécessaire. On se place alors dans un cas différent où aucune annotation relative à la tâche n'existe. Il s'agit d'un apprentissage dit « non-supervisé » dont l'objectif est d'analyser la distribution statistiques des données dans l'espace engendré par les caractéristiques extraites. Dans cet espace, on cherche à isoler des groupes de données proches relativement à une métrique. Ainsi, trois problèmes essentiels se posent :

1) Quel algorithme utiliser pour découvrir automatiquement des groupes similaires dans un ensemble non-structuré de données ;
2) Quelles caractéristiques extraire des données de manières à ce que des groupes distincts se créent ;
3) Quelle métrique utiliser pour que ces groupes soient le plus disjoints possibles.

Ici encore, il n'existe pas de réponse universelle à ces problèmes. Certains algorithmes nécessitent la connaissance *a priori* du nombre de groupe recherchés, d'autres impliquent la projection des données dans des espaces de représentation différents. Dans le projet METHODEO, il s'agit en plus de définir des groupes de données sur lesquelles les algorithmes auront des performances similaires.

Conclusion

Le projet METHODEO a comme objectif de définir une méthodologie générique d'évaluation des algorithmes d'exploitation des enregistrements dans le cadre des enquêtes judiciaires L'idée principale est d'aider les opérateurs à prévoir avec une bonne précision le comportement des algorithmes qui accomplissent une certaine fonctionnalité sur des données réelles (dans un environnement réel) en faisant une comparaison entre la description des nouvelles données avec les descriptions des classes de vidéos sur lesquelles les algorithmes ont déjà été évalués.

Pour cela les vidéos sont décrites de la façon la plus précise, exhaustive et objective possible. Cette description contient des descripteurs vidéo de bas niveau et des éléments (texte, boites englobantes) décrivant le contexte de la scène et le scenario (les objets et les événements). Ces éléments sont structurés dans un format basé sur la norme ISO 22311. Beaucoup de questions n'ont pas trouvé encore des réponses universelles ce qui donne lieu à beaucoup de perspectives. Le défi du cout de la génération des annotations a attiré des équipes de recherche qui travaillent actuellement sur les aspects de semi automatisation des annotations [4]. Un autre grand domaine de recherche est la gestion de la mobilité des caméras, des objets d'intérêt et des utilisateurs dans le cadre d'un réseau de caméras [5].

Références

[1] V. Vapnik, *The nature of statistical learning*, Springer-Verlag, 1995

[2] V.T. de Almeida, R. Gutting, *Indexing the trajectories of moving objects in networks.* GeoInformatica 9(1), 30-36, 2005.

[3] Glenn Hartmann et al, *Weakly Supervised Learning of Object Segmentations from Web-Scale Videos.* ECCV 2012.

[4] C. Vondrick, D. Ramanan, *Video Annotation and Tracking with Active Learning*, NIPS 2011.

[5] D. Petrovska et al, *Guide to Biometric Reference Systems and Performance Evaluation.* Springer-Verlag 2009