

RAPPORT
Chef d'Oeuvre
Méthodes et Algorithmes

COURDY-BAHSOUN CLÉMENCE - DEKER SYLVAIN
KOTTATH SANDEEP - MOUSSA NAHOR - YAZI RYMA

M2 IGAI

23.11.2019

1 Introduction

Le terrorisme et la criminalité sont des problématiques d'envergure mondiale, et dans cette lutte menée par les forces de l'ordre de nombreuses recherches dans la vision par ordinateur sont menées pour faciliter la ré-identification de personne ou de véhicule dans des bandes vidéo pouvant être longues et pas toujours de bonne qualité. Dans ce contexte le projet proposé par le client, Mr Alain Crouzil, est le développement d'un outil de ré-identification de véhicule dans des enregistrements. Plus précisément il s'agit de retrouver dans un ensemble de vidéo le véhicule correspondant à celui recherché. Ainsi l'objectif est de faire un éditeur permettant de retrouver un véhicule dont l'image est donnée en paramètre dans des vidéos de surveillances fournit par une base de donnée. Pour ce faire il est imposée d'utiliser une approche de résolution par apprentissage profond en utilisant des réseaux de neurones existant comme le propose les détecteurs d'objets de type YOLO et SSD. Après quoi en mesurant la différence entre les résultats obtenus en sortie du réseau pour l'image de référence (véhicule recherché) et ceux obtenus pour les frames des enregistrements on pourra trier par pertinence ces différences, et ainsi afficher la liste des résultats les plus probants.

La réalisation de ce projet se base sur divers articles scientifiques traitant de la ré-identification et présentant diverses approches et divers outils pouvant être utilisés. Ainsi par l'analyse de ces articles deux méthodes sont mises en évidence par leurs résultats positifs aussi bien en taux de ré-identification qu'en temps de calcul. Il s'agit des méthodes Fusion of Attributes and Color features (FACT) et SSD toutes deux basées sur l'exploitation de réseaux de neurones profonds mais par des approches différentes. Pour rappel un réseau de neurones et un ensemble de neurones connectés entre eux et qui permettent de résoudre des problèmes complexes tel que la reconnaissance de forme, ou ici la ré-identification à l'issue d'un entraînement connu sous le nom de phase d'apprentissage. Un neurone, au sens informatique du terme, est un dispositif prenant en entrée plusieurs paramètres et fournissant un seul paramètre de sortie et qui est modélisé pour reproduire certaines propriétés d'un neurone biologique. Si l'approche nécessite l'entraînement du, ou des réseaux utilisés, il existe des bases d'apprentissage riches et varié comme CompCars ou Veri qui sont accessible en open source.

La partie interface du logiciel n'est pas présentée ici, il est tout de même à noter qu'elle sera faite en python en exploitant PyQt. Pour le traitement des bandes vidéos la librairie opencv sera utilisée.

Une analyse des méthodes FACT et SSD a permis de définir l'approche de résolution du problème de ré-identification à mettre en oeuvre qui est présentée.

2 Méthodes et algorithmes

Les méthodes de détection sont toutes basées sur des réseaux de neurones. Chaque méthode se base sur des caractéristiques différentes : la couleur, la texture ou la sémantique c'est à dire la forme, les détails.

2.1 FACT

La méthode Fusion of Attributes and Color feaTures (FACT) consiste à fusionner plusieurs méthodes d'identification. Les descripteurs utilisés sont :

1. le **BOW-SIFT (Bag Of Words Scale-Invariant Feature Transform)** basé sur la texture. Il permet d'extraire le descripteur SIFT qui est une méthode de mise en correspondance de point utilisé en reconnaissance de forme à partir du modèle BOW qui est implémenté par une approche des k-moyenne hiérarchique.
2. le **BOW-CN (Color Name)** basé sur la couleur. BOW-CN est également construit sur le modèle BOW duquel on extrait le descripteur de couleur en tenant compte de contrainte géométrique.
3. le réseaux de neurones convolutif **GoogleLeNet** basé sur la sémantique.

Le principe est le suivant : A partir des scores obtenus par une distance euclidienne entre le résultat et la vérité terrain pour les trois descripteurs obtenus on effectue une somme pondérée, avec poids fixes définis par expérimentations. La valeur des poids est proportionnel à l'importance du descripteur pour le protocole de ré-identification. Ainsi la pondération sera plus forte sur le descripteur de la sémantique qui permet d'identifier les détails, et moindre sur le descripteur de texture et de couleur. D'après les tests menés sur la base de donnée Veri, FACT offre un taux de réussite d'environ 75% en un temps d'exécution bien meilleure que les différentes méthodes de même catégories testées. Cependant il est important de noter qu'à ce jour une des causes d'échec de la méthode FACT est liée à un problème d'éclairage. En effet en fonction de l'éclairage la couleur d'un véhicule peut être perçu différemment, ce qui peut fausser le résultat. En dépit de cela cette méthode présente de nombreux avantages, et les recherches actuelles tendent à résoudre ce problème de couleur lié à l'éclairage.

2.2 SSD

L'approche SSD est basée sur l'ajout de plusieurs niveaux de convolution, de plus en plus petit, à la fin du réseau de neurones. Cette méthode utilise également

des boites englobantes, à définir par défaut, pour chaque caractéristiques ce qui va faciliter la détection de caractéristiques particulières. Les expérimentations menées montre un taux de réussite d'environ 76% pour un très bon temps de calcul. Cependant cette méthode n'est pas adapter pour notre étude car elle est basée sur la détection de différences majeure tel qu'il existe par exemple entre un humain et une voiture. Hors ré-identification de véhicule nécessite des comparaisons très précises sur des modèles de véhicules. De plus des tests réalisés on démontrés une erreur d'environ 35% sur les véhicules c'est à dire que 35% des véhicules trouver dans les images n'en sont pas, ce qui est clairement un soucis dans le cadre de la ré-identification de véhicule.

2.3 Approche de résolution

Le projet se divise en plusieurs parties : l'interface , la méthode d'exploitation par apprentissage profond et le traitement des résultats.

L'interface sera réalisée en PyQt, un outils permettant de faire des interface en Python. Elle devra permettre à l'utilisateur de rechercher un véhicule dans un ensemble d'enregistrements.

Les réseaux de neurones peuvent être récupérés en étant déjà entraînés ou bien nous pourrons nous même les exercer mais cela risque de prendre du temps. Il est possible avec la méthode FACT de partir de réseau ayant déjà subit la phase d'apprentissage pour le problème qui nous intéresse. Ainsi nous opterons dans la mesure du possible pour des réseaux pré-entraîné, grâce à des bases d'apprentissage telles que CompCars et Veri qui sont riches en information, et prêt à être utilisé pour la détection de véhicule.

Les résultats seront les vidéos ainsi que l'heure d'apparition du véhicule. Les frames étant à la suite si le véhicule est détecté dans une frame f nous ignorerons les frames des 60 secondes suivantes. Pour chaque détection nous stockerons son score afin de pouvoir afficher ces derniers par ordre croissant. De cette manière les résultats les plus vraisemblable seront les premiers résultats. Le score sera a priori calculé par un simple calcul de distance euclidienne entre le résultat obtenu pour l'image de référence et les vidéo où un véhicule semblable a été repéré, donnant ainsi une mesure de la dissemblance : plus le score est petit et plus la ressemblance est grande. Cependant en fonction des résultats obtenus, on pourra éventuellement tester d'autres méthodes de comparaison, basées sur d'autres normes comme la mesures de corrélation à partir de la norme L_p . Ou encore une mesure de corrélation croisée dont les résultats sont normalisés entre 0 et 1. A priori il ne semble pas nécessaire d'aller aussi loin car l'objectif est de classer les résultats, on peut alors tout simplement se contenter du résultat de la distance euclidienne au carré,

évitant ainsi le calcul d'une racine carré non nécessaire à la précision du résultat.

3 Conclusion

Pour résoudre le problème de la ré-identification de véhicule dans des enregistrements de nombreuses méthodes ont vu le jour. Parmi elles deux se démarquent particulièrement : la méthode FACT qui présente d'excellents résultats en qualité et en temps par rapport aux autres méthodes de sa catégorie, et la méthode SSD qui malgré de bons résultats semblent être plus adaptés pour identifier les différences entre deux objets. Cependant l'objectif étant de définir une liste triée par pertinence des vidéos pouvant contenir le véhicule recherché, nous avons décidé d'opter pour la méthode FACT pour sa capacité à détecter et traiter des détails.

Pour ce qui est du tri des vidéos par pertinence, dans chaque vidéo identifiée, les résultats entre l'image du véhicule recherché et celle de la première frame voyant apparaître le véhicule seront comparés par un simple calcul de distance euclidienne au carré, évitant ainsi de s'encombrer d'un calcul de racine peu utile pour le but recherché. Plus la distance calculée sera petite et plus il y aura de chance que le véhicule reconnu sur la vidéo soit le même que celui recherché.

Les risques à prendre en compte concernent les mauvaises conditions d'enregistrement ou de capture : avoir un angle de vue différent par rapport à celui de l'image par exemple. De plus nous ne sommes pas à l'abri de potentiels soucis d'intégration ou de fonctionnement, lors de l'entraînement des réseaux (si nécessaire) il nous faudra disposer d'une grande puissance de calcul.