

Sujet de thèse

Titre de la thèse

Techniques d'apprentissage pour la configuration automatique du processus d'interprétation de scènes.

Mots-clefs

Apprentissage supervisé/non supervisé, évaluation des performances, interprétation de séquences d'images, classification, suivi, reconnaissance de scénarios, système autonome auto-adaptatif.

Contexte du sujet

Cette thèse se situe dans le cadre des travaux du projet ORION sur l'interprétation automatique de scènes pour la reconnaissance de comportements humains.

L'équipe ORION a conçu une plate-forme d'interprétation de séquences vidéo (appelée VSIP) permettant la détection d'objets mobiles, le suivi de leur trajectoire et la reconnaissance de certains comportements prédéfinis par des experts humains dans une base de scénarios. Cette plate-forme comporte un grand nombre de paramètres à régler, plus d'une centaine (par exemple les paramètres de la segmentation ou du suivi) et de connaissances à configurer (modèle géométrique des objets attendus dans la scène, modèle de la scène vide) afin d'obtenir une interprétation pertinente de la scène observée. Cette plate-forme a été notamment testée sur des applications de vidéosurveillance de métros, de banques et de maintien de personnes âgées à domicile.

La perception du monde réel est dynamique et incertaine. Ainsi, les processus d'interprétation d'images sont souvent erronés mais difficiles à évaluer. Le programme Techno-Vision ETISEO (janvier 2005-2007) a pour objectif d'évaluer les techniques d'interprétation vidéo appliquées à la vidéo surveillance. Le projet ORION a proposé des métriques d'évaluation, un outil de comparaison automatique des résultats d'analyse vidéo avec des vérités terrain. Ce programme d'évaluation a montré la nécessité d'automatiser la configuration des systèmes interprétation de vidéos créés à partir de la plate-forme VSIP ainsi que le réglage dynamique des paramètres des systèmes afin de tenir compte des modifications et évolutions de l'environnement de la scène (e.g. changement de conditions d'illumination).

Description du sujet

L'objectif de cette thèse est de proposer des techniques d'apprentissage afin d'obtenir un processus d'interprétation de scènes, autonome, capable de se configurer automatiquement, de s'adapter à tout changement et évolution de environnement, de fonctionner en mode dégradé et d'accumuler des connaissances sur son fonctionnement pour s'améliorer en cours d'utilisation.

Une première étape consistera à étudier globalement le processus d'interprétation de scène et de qualifier la performance des résultats obtenus. Plus exactement, il s'agit de comprendre et de définir les conditions d'utilisation des algorithmes constitutifs de la plate-forme VSIP et toute la connaissance a priori nécessaire pour leur bon fonctionnement (calibration pour passer du monde capteurs au monde 3D, position des capteurs, géométrie de la scène, disposition de l'éclairage, typologie des objets mobiles évoluant dans la scène et leurs comportements associés,...). Au cours de cette thèse, on s'attachera à classer les données d'entrée des algorithmes (e.g. les vidéos, les objets mobiles), les objectifs à attendre (e.g. segmentation en régions de mouvement, événements pre-définis par l'utilisateur), et les paramètres des algorithmes afin de prévoir la qualité des résultats possibles d'interprétation.

Une deuxième étape étudiera différentes techniques d'apprentissage supervisées afin d'apprendre automatiquement la valeur des paramètres des algorithmes constitutifs de la plate-forme VSIP en fonction des caractéristiques des flux vidéos acquis par les caméras. Dans un premier temps, on utilisera la classification des données d'entrée des algorithmes réalisée lors de la première étape et on étudiera les techniques classiques d'apprentissage supervisées (par exemple, méthode d'exploration d'un espace de paramètres telle que la méthode du simplexe). Pour cela, on dispose d'un ensemble de vérités terrain fournies par le programme ETISEO décrivant les caractéristiques des objets mobiles, leurs suivis, et les événements d'intérêt pour un nombre important de séquences vidéos représentatives d'une grande diversité d'environnements (extérieur/intérieur, individus isolés/foule, personnes/vélos/voitures/trains, ombres/pénombre/lumineux, mono-caméra/multi-caméras). L'étape d'apprentissage consiste à relier ces caractéristiques des données d'entrée aux valeurs des paramètres des algorithmes.

Une troisième étape consistera à étudier les capacités d'autonomie du processus d'interprétation de scènes. En effet, lorsque les performances se dégradent, il peut être nécessaire de prendre en compte de nouvelles caractéristiques des données d'entrée (non représentées dans la base initiale d'apprentissage) étant donné leur grande variabilité (e.g. une situation de foule succédant à une situation de scène vide) et leur grande diversité, et de doter le système d'interprétation de capacité d'apprentissage de ces nouvelles caractéristiques avec une faible aide extérieure (faible intervention d'un opérateur humain). Pour cela, on étudiera dans un deuxième temps des techniques d'apprentissage incrémental bénéficiant des travaux précédents sur les techniques supervisées. Le processus d'interprétation rassemblera les nouvelles caractéristiques des données d'entrée

dans de nouvelles catégories lors d'un traitement sur une longue période (par exemple, une semaine). Lorsqu'une catégorie devient suffisamment importante, le système montre à l'opérateur humain cette nouvelle catégorie et lui demande par exemple de donner les vérités terrain associées aux résultats attendus. Il s'agira donc d'un apprentissage incrémental faiblement supervisé. Une alternative aux vérités terrain consistera pour le développeur à proposer une modification des algorithmes. Enfin, on proposera des techniques de pilotage de programmes afin d'automatiser la réparation des algorithmes.

Les techniques d'apprentissage proposées dans la thèse seront validées dans le cadre du projet SIC (Sécurité des Infrastructures Critiques) du pôle de compétitivité de la région Paris Sud. Lors de ce projet, il s'agira d'utiliser la plate-forme VSIP afin d'évaluer les performances de la vidéo surveillance automatique pour résoudre des problèmes de sécurité dans plusieurs contextes applicatifs (sécurisation de grands événements tels que des matchs de football, sécurisation de bâtiments publiques tels que la Mairie de Bobigny et sécurisation des gares et transports publiques).

Contributions attendues

- Une nouvelle méthode d'apprentissage incrémental faiblement supervisé pour l'initialisation des paramètres
- Une nouvelle méthode d'adaptation automatique d'un processus d'interprétation en cas de changement des conditions de fonctionnement.
- Validation des techniques d'apprentissage à travers des applications de vidéo surveillance.
- Publications dans des revues et conférences internationales telles que ECCV (European Conference on Computer Vision) , ICCV (International Conference on Computer Vision), PAMI (Pattern Analysis and Machine Intelligence) ou ECAI (European Conference on Artificial Intelligence).

Planning

Première année :

- Etat de l'art sur l'interprétation de scènes, évaluation des performances et les techniques d'apprentissage.
- Etude globale du processus d'interprétation de scènes de la plate-forme VSIP et classification des données d'entrée et des paramètres des algorithmes de la plate-forme.

Deuxième année :

- Apprentissage supervisé des classes des paramètres des algorithmes à partir de vérités terrain.
- Validation des techniques d'apprentissage à l'aide du corpus ETISEO.

Troisième année :

- Analyse du comportement du processus d'interprétation en mode dégradé (capteurs en panne,...)
- Apprentissage incrémental faiblement supervisé de nouvelles classes de paramètres.
- Réparation automatique du processus interprétation

Références

1. F. Cupillard, A. Avanzi, F. Brémont and M. Thonnat, Video Understanding for Metro Surveillance. The IEEE ICNSC 2004 in the special session on Intelligent Transportation Systems, Taiwan, March 2004.
2. T. Vu, F. Brémont and M. Thonnat, Automatic Video Interpretation: A Novel Algorithm for Temporal Scenario Recognition. The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03), Acapulco, Mexico, 9-15 August 2003.
3. Alberto Avanzi, Francois Bremond, Christophe Tornieri and Monique Thonnat, Design and Assessment of an Intelligent Activity Monitoring Platform, in EURASIP Journal on Applied Signal Processing, special issue in "Advances in Intelligent Vision Systems: Methods and Applications", 2005.
4. Van-Thanh VU. Temporal Scenarios for Automatic Video Interpretation. Université de Nice – Sophia Antipolis, France, 2004
5. B. Georis, Program Supervision Techniques for Easy Configuration of Video Understanding Systems. PhD thesis, INRIA – Université Catholique de Louvain, Belgique, 2006.
6. F. Brémont, Environnement de résolution de problèmes pour l'interprétation de séquences d'images. PhD thesis, INRIA - Université de Nice Sophia-Antipolis, 1997.

Lieu de la thèse

- INRIA Sophia-Antipolis, projet Orion.

Directeur de thèse

Monique THONNAT, HDR et DR INRIA
Projet ORION, INRIA-Sophia Antipolis
2004 route des Lucioles BP 93
06902 Sophia Antipolis Cedex
Tél. : 04 92 38 78 67
Email: Monique.Thonnat@sophia.inria.fr

Co-encadrant

François Bremond, CRI INRIA
Projet ORION, INRIA-Sophia Antipolis
2004 route des Lucioles BP 93
06902 Sophia Antipolis Cedex
Tél. : 04 92 38 76 59
Email: Francois.Bremond@sophia.inria.fr