

ED SMAER

Sujet de thèses 2014

Laboratoire : Institut des Systèmes Intelligents et de Robotique

Etablissement de rattachement : UPMC Sorbonne Universités

Directeur de thèse et section CNU ou CNRS : Catherine Achard, section 61

Codirection et section CNU ou CNRS : Séverine Dubuisson, sections 27-61

Titre de la thèse : Analyse d'anomalies dans les interactions sociales. Application à la détection de comportements humains suspects ou dangereux dans les séquences vidéo

Collaborations dans le cadre de la thèse :

Rattachement à un programme :

Le sujet peut être publié sur le site web de l'ED SMAER : OUI ~~NON~~ rayer la mention inutile.

Résumé du sujet :

NOTA : l'ED SMAER collecte environ 100 fiches par an. Merci de limiter cette fiche à 1 page maxi.

La détection de comportements suspects voire dangereux est devenue d'intérêt majeur pour notre société. La multiplication des caméras dans les lieux publics résulte dans une explosion du nombre d'écrans dans les centres de contrôle, rendant la vidéosurveillance humaine difficilement réalisable. Il devient donc indispensable de développer des systèmes automatiques de détection d'événements anormaux qui se déroulent dans des lieux publics tels que les métros, les places, les centres commerciaux, les files d'attente dans les magasins, *etc.*

Cette thèse a pour objectif de proposer une approche originale d'analyse de comportements suspects puisqu'elle abordera le problème du point de vue de l'analyse des interactions sociales. En effet, un comportement dangereux, tel qu'une agression par exemple, se caractérisera comme une rupture lors d'une interaction sociale. Le contexte applicatif de cette thèse concernera la détection de comportements suspects, tels qu'une agression, qu'elle soit physique ou verbale, ou encore un comportement atypique vis-à-vis des autres, *etc.* Il sera ainsi envisagé la création d'une base spécifique dans laquelle des scénarios d'interaction violente seront simulés, puis une validation dans des conditions de vidéo-surveillance

Sujet développé (à présenter en 2 ou 3 pages maximum,
en précisant notamment le contexte, les objectifs, les résultats attendus)

Analyse d'anomalies dans les interactions sociales. Application à la détection de comportements humains suspects ou dangereux dans les séquences vidéo

Contexte :

La détection de comportements suspects voire dangereux est devenue d'intérêt majeur pour notre société. La multiplication des caméras dans les lieux publics résulte dans une explosion du nombre d'écrans dans les centres de contrôle, rendant la vidéosurveillance humaine difficilement réalisable. Il devient donc indispensable de développer des systèmes automatiques de détection d'événements anormaux qui se déroulent dans des lieux publics tels que les métro, les places, les centres commerciaux, les files d'attente dans les magasins, *etc.*

Objectifs :

Cette thèse a pour objectif de proposer une approche originale d'analyse de comportements suspects puisqu'elle abordera le problème du point de vue de l'analyse des interactions sociales. En effet, un comportement dangereux, tel qu'une agression par exemple, se caractérisera comme une rupture lors d'une interaction sociale. Le but est donc triple : proposer un ou des descripteurs locaux ou semi-locaux permettant de caractériser l'interaction, puis modéliser l'interaction et enfin détecter le caractère « anormal » d'une interaction. Ces trois objectifs sont brièvement décrits par la suite.

1) Dans un premier temps, il faudra, par l'analyse des langages corporels et audio, définir *via* des descripteurs compacts locaux et globaux l'essence d'une interaction sociale (par exemple, savoir caractériser de manière simple et résumée un échange entre deux ou plusieurs personnes ou encore un comportement individuel vis-à-vis d'un groupe tel qu'il est de manière « traditionnelle »). Ces descripteurs devront intégrer à la fois des composantes descriptives (spatiales) et dynamiques (temporelles). De nombreux travaux récents ont proposé de nouveaux descripteurs d'images permettant de détecter rapidement des caractéristiques importantes dans les images [LUO-12, CHA-12, AMB-11, RUB-11, LAP-05], ou encore des méthodes robustes de détection d'objets mobiles dans les séquences vidéo [CHO-12]. De même, du fait des variations de contenu qui peuvent intervenir dans ces séquences (arrivée et départ d'acteur, occultations, changements colorimétriques, *etc.*), nous nous focaliserons sur les descripteurs locaux et semi-locaux. Ces descripteurs seront étudiés au vu de l'état de l'art actuel, ce qui permettra d'identifier les plus appropriés à nos besoins, parmi les points d'intérêt, régions, contours locaux *edgels*), *etc.*

2) Ensuite, il faudra modéliser l'interaction, telle qu'elle est perçue par les différents acteurs de la scène, ce qui permettra de mieux en comprendre les mécanismes (*i.e.* de l'action - cause - à la réaction - conséquence -). Ici il sera envisagé un modèle permettant une interprétation (ou caractérisation de l'interaction), qui intègre un certain nombre de signaux sociaux non verbaux (gestes, posture, voire même le rythme de la parole) ainsi que le ou les descripteurs qui auront été développés dans la partie 1. Là encore, l'objectif est de proposer un modèle

compact et descriptif de l'interaction qui soit le plus général possible. En effet, c'est à partir de ce modèle qu'on tâchera de reconnaître une interaction spécifique, ce qui est le travail de la partie 3.

3) Enfin, il faudra fournir une mesure qui permettra de détecter une rupture entre la description d'une interaction « normale » et une interaction « suspecte ». En particulier, il faudra donc répondre à la question : est-ce que l'interaction observée correspond à un échange « normal » entre deux ou plusieurs personnes et, si non, en quoi semble-t-elle différente ? A partir des modèles développés dans la partie 2, on définira des espaces d'observation qui permettront d'en définir des caractéristiques. On identifiera bien le comportement de chaque acteur de la scène pour essayer de globaliser l'analyse de l'action/réaction. Nous souhaitons identifier dans ces espaces de représentation des groupes de comportements dans les interactions qui permettront de reconnaître un type d'interaction.

Résultats attendus :

Le contexte applicatif de cette thèse concernera la détection de comportements suspects, tels qu'une agression, qu'elle soit physique ou verbale, ou encore un comportement atypique vis-à-vis des autres, *etc.* La chaîne de traitement, telle qu'elle est définie plus haut sera validée au fur et à mesure de son élaboration : des descripteurs à la reconnaissance d'interactions anormales, voire violentes. Il existe de nombreuses bases qui nous permettront de tester localement la pertinence du modèle, même en dehors du contexte applicatif. Il sera ainsi envisagé la création d'une base spécifique dans laquelle des scénarios d'interaction violents seront simulés. Cette base permettra de tester les modèles sur des données relativement propres, et de les valider à l'aide d'une vérité terrain que nous ferons nous-même. Ces modèles seront ensuite validés dans des conditions de vidéo-surveillance (avec éventuellement plusieurs caméras) et pour cela des bases de données existantes seront utilisées (UCSD Anomaly Detection Dataset [UCSD], Boss Dataset [BOSS]), ainsi que d'autres plus centrées sur les comportements criminels ([PETS14]) tels que le vol ou les agressions.

Bibliographie

- [UCSD] UCSD Anomaly Detection Dataset. <http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm>
- [BOSS] Boss Dataset. <http://www.multitel.be/image/research-development/research-projects/boss.php>.
- [PETS14] PETS 2014 dataset. <http://www.cvg.rdg.ac.uk/PETS2014/a.html>.
- [LAP-05] I. Laptev. On space-time interest points. *International Journal of Computer Vision*, 2005, 64 (2-3):107-123.
- [AMB-11] M. Ambai and Y. Yoshida. CARD : compact and real-time descriptors. *IEEE International Conference on Computer Vision*, 2011, pp. 97-104
- [CHA-12] B. Chakraborty, M. B. Holte, T. B. Moeslund and J. Gonzalez. Selective spatio-temporal interest points. *Computer Vision and Image Understanding*, 2012, 116(3) :396-410.
- [LUO-12] W. Luo, H. Lia, G. Liua and K. Ngi Nganb. Global salient information maximization for saliency detection. *Signal Processing: Image Communication*, 2012, available online.
- [RUB-11] E. Rublee, V. Rabaud, K. Konolige and G. Bradski. ORB : an efficient alternative to SIFT and SURF. *IEEE International Conference on Computer Vision*, 2011, pp. 2564-2571.