

Stage fin d'études ou master:

Analyse des foules en utilisant des caractéristiques automatiquement apprises par des auto-encodeurs convolutionnels (Deep learning)

Présentation du sujet :

Avec la propagation des caméras de la vidéo surveillance, les techniques de la vision par ordinateur et d'apprentissage jouent un rôle primordial pour analyser les vidéos. Précisément, il s'agit de faire appel à certains algorithmes qui permettent d'analyser automatiquement la scène au lieu de la superviser manuellement par un agent de sécurité.

Bien que la majorité des approches actuelles en vidéo surveillance supposent que les scènes sont simples et très peu denses, certains événements ou des expositions de grande envergure nous rend la tâche plus difficile avec des scènes beaucoup plus complexes (et surtout denses). Pour cette raison, on constate qu'aujourd'hui l'analyse des foules [1,2] occupe plus d'importance pour pouvoir les contrôler et les gérer, notamment pour des raisons de sécurité. Cette analyse peut être utile pour anticiper et empêcher les débordements potentiellement dangereux, surtout lorsque le nombre de personnes dépasse certaines limites. Outre son utilité dans le domaine de la sécurité, l'analyse des foules s'avère pertinente aussi pour des applications dans le domaine économique comme l'organisation des emplois du temps des personnels dans les centres commerciaux et les horaires des services de transport public.

Pour résoudre les problèmes inhérents à la gestion des foules, des nombreux travaux [2] ont été réalisés, dans les quels, la représentation de la scène se fait toujours par des caractéristiques bas ou moyen niveau telles que les tracklets, les points d'intérêts, des caractéristiques de la texture [3,6]...

Ce travail a pour objectifs, d'apporter des solutions à l'analyse des foules. Notamment, le stage consiste en la conception, le développement et l'évaluation de méthodes issues du Deep Learning [5] appliquées à l'analyse des foules (la vidéo surveillance au sens large du terme).

Par rapport aux solutions existantes, ce travail nécessite une réflexion sur la représentation des scènes en utilisant des bibliothèques de Deep Learning. Il s'agit particulièrement de la caractérisation des images de la foule par des caractéristiques automatiquement apprises par des auto-encodeurs convolutionnels [4, 5]. La pertinence de ces caractéristiques apprises avec des réseaux de neurones profonds doit être démontrée en menant les mêmes tests avec d'autres caractéristiques conventionnelles (LBP, HOG, KLT ...).

Travail demandé :

C'est un travail de recherche et développement qui a pour objectifs :

- de résoudre le problème de la caractérisation des scènes denses (les foules) par des auto-encodeurs convolutionnels.
- d'appliquer cette caractérisation dans une application de l'analyse des foules (estimation de la densité, reconnaissance d'événement ou détection d'anomalie)
- de démontrer que cette caractérisation automatique de la scène peut considérablement améliorer les résultats.

Pour conclure, il s'agit d'un travail de recherche et développement ayant comme objectif d'explorer un domaine d'étude à la fois prometteur et stimulant. C'est un travail qui vise une application intéressante (l'analyse des foules) avec une approche innovante (l'apprentissage profond).

Environnement technique :

C++/JAVA ou Python comme langage de programmation avec la bibliothèque **OpenCV**

Mots-clés:

Vision par ordinateur/ traitement d'images, apprentissage automatique (machine learning), intelligence artificielle, analyse des foules, Deep learning

Lieu :

Laboratoire LATIS à l'ENISO ou ISSAT de Sousse

Remarque:

Ce travail peut être réalisé dans le cadre d'un **PEF** ou d'un **stage master M2**. Dans les deux cas il y a une possibilité de poursuite en **thèse**. Le travail demandé (entre la recherche et le développement) sera adapté en fonction du profil de candidat (élève ingénieur ou M2).

Profil du candidat :

Ce travail nécessite un candidat avec :

- Formation Master 2 Recherche et/ou dernière année formation ingénieur informatique
- Compétences requises : un bon niveau de programmation en C++, ou JAVA ou Python
- Compétences souhaitées : bonnes connaissances en vision par ordinateur, traitement d'images, machine learning (apprentissage statistique) seront appréciées.

- Un bon niveau d'anglais est nécessaire

Le candidat devra être capable de rechercher et proposer des solutions sur des problèmes tant techniques que scientifiques. Il doit avoir un fort intérêt pour la technologie et la lecture scientifique et technique et une capacité continue à apprendre, s'adapter aux nouveautés et évoluer au fil du temps.

Contact:

Hajer Fradi (fradi.hajer.issat@gmail.com)

Modalités de candidature :

Adresser un CV détaillé à l'adresse suivante : fradi.hajer.issat@gmail.com

Références bibliographiques :

[1] J. Jacques, Junior, S. Raupp Musse, and C. Jung. Crowd analysis using computer vision techniques. *IEEE Signal Processing Magazine*, 27(5):66–77, 2010.

[2] H. Fradi. New insights into crowd density analysis in video surveillance systems. *Image Processing*. Télécom ParisTech, 2014

[3] J. Shao, C. C. Loy, and X. Wang Scene-Independent Group Profiling in Crowd, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[4] J. Shao, K. Kang, C. C. Loy, and X. Wang, Deeply Learned Attributes for Crowded Scene Understanding, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[5] Jürgen Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks*, Volume 61, Pages 85-117, 2015

[6] H. Fradi, B. Luvison, and Q. C. Pham. "Crowd behavior analysis using local mid-level visual descriptors. *IEEE Transactions on Circuits and Systems for Video Technology* 27 (3), 589-602, 2017