CEA List
Laboratoire de Vision et d'Ingénierie des Contenus
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contact
Tél +33 (0)1 69 08 01 37
E-mail bertrand.luvison@cea.fr

STAGE 2019 Réf : AS-19-S3

Estimation de pose 3D humaine par deep learning

Présentation du laboratoire d'accueil

Au cœur du Plateau de Saclay (Ile-de-France), l'institut CEA LIST focalise ses recherches sur les systèmes numériques intelligents. Porteurs d'enjeux économiques et sociétaux majeurs, ses programmes de R&D sont centrés sur les systèmes interactifs (intelligence ambiante), les systèmes embarqués (architectures, ingénierie logicielle et systèmes), les capteurs et le traitement du signal (contrôle industriel, santé, sécurité, métrologie).

Dédiés à la recherche technologique, les 700 ingénieurs-chercheurs et techniciens de l'institut ont pour objectif de favoriser l'innovation et son transfert autour de partenariats industriels pérennes. La culture projet et l'excellence scientifique des équipes de l'Institut sont au cœur de cette ambition.

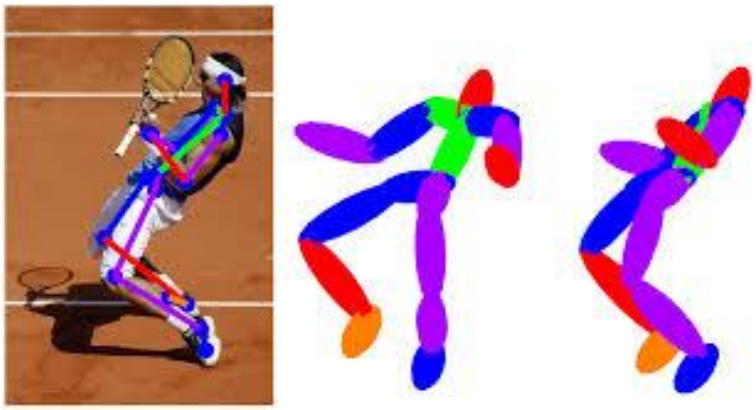
Le Laboratoire de Vision et d'Ingénierie des Contenus mène ses recherches dans les domaines de la Vision par Ordinateur (Computer Vision) et l'analyse automatique de texte avec le défi d'extraire et d'organiser l'information à partir de documents faiblement ou non structurés (texte, image, vidéo, réseaux de capteurs).

Contexte du stage

Ce stage s'inscrit dans les activités d'Analyse de Scène du Laboratoire Vision et Ingénierie des Contenus du CEA List. Le laboratoire développe dans le cadre de projets R&D, des systèmes d'analyse automatique d'images et de vidéo intégrant des technologies avancées de détection, localisation, classification et suivi temporel d'objets d'intérêt, de reconnaissance d'activité et de comportements, ainsi que des outils de modélisation automatique de la scène.

Description du stage

Comprendre automatiquement l'activité humaine se déroulant dans un flux vidéo est l'un des principaux challenges actuels de la vision par ordinateur. Bien que le sujet soit abordé depuis longtemps, les applications qui en résultent ne se limitent qu'à des contextes bien particuliers avec des gestes bien particuliers [1]. L'un des prérequis à cette analyse est bien souvent l'extraction d'une information de type posture des personnes présentes de la scène, afin d'avoir une représentation beaucoup plus compacte de l'état des personnes à un instant t. C'est d'ailleurs lorsque ce type d'information a pu être extrait de manière fiable, notamment avec des capteurs tel que la Kinect, que de nettes avancées ont pu être réalisées [2]. Les progrès notables de ces dernières années, ont notamment été réalisé pour l'estimation de squelette 2D à partir d'images RGB uniquement [3, 4] mais aussi pour l'estimation de squelette 3D à partir d'image RGB [5, 6]. Sur ce dernier point l'estimation de squelette 3D souffre encore de nombreux problèmes, tel que la robustesse face au pose inhabituelle, la robustesse dans des conditions « in the wild », c'est-à-dire dans des environnements quelconques et variés, la gestion du multi personne, etc.



Estimation de pose 3D à partir d'une image RGB

CEA List
Laboratoire de Vision et d'Ingénierie des Contenus
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contact
Tél +33 (0)1 69 08 01 37
E-mail bertrand.luvison@cea.fr

L'objectif du stage sera donc de s'attaquer à ces faiblesses en s'articulant autour de deux phases. La première consistera à mettre en place une plateforme d'acquisition basée sur des capteurs de type Kinect ou de « Motion Capture », permettant d'annoter facilement les images en mimant les postures 3D de personnes issues d'images « in the wild » qui ne sont pour le moment annotées qu'en 2D (COCO keypoint, MPII 2D dataset, etc). La seconde phase, consistera à exploiter ces données générées pour l'estimation de squelettes 3D sur ces images à l'aide de réseau de neurones convolutionnels qu'il faudra définir durant le stage.

Références:

[1] Information Fusion for Action Recognition with Deeply Optimised Hough Transform Paradigm. Vaquette, G., Achard, C., & Lucat, L. (2016). In VISIGRAPP (4: VISAPP) (pp. 423-430).

[2] Real-Time Skeleton-Tracking-Based Human Action Recognition Using Kinect Data. G.Papadopoulos, A. Axenopoulos, P. Daras, in Computer Science, Springer, 2014

[3] Realtime multi-person 2d pose estimation using part affinity fields. Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2016), In CVPR 2016

[4] Mask R-CNN. Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, In CVPR 2017

[5] Towards 3d human pose estimation in the wild: a weakly-supervised approach. Zhou, X., Huang, Q., Sun, X., Xue, X., & Wei, Y. (2017, October). In IEEE International Conference on Computer Vision.

[6] Ordinal depth supervision for 3d human pose estimation. Pavlakos, G., Zhou, X., & Daniilidis, K. (2018). In CVPR 2018,

Niveau demandé :	Ingénieur, Master 2
Ce stage ouvre la possibilité de poursuite en thèse et ingénieur R&D dans notre laboratoire.	
Durée :	6 mois
Rémunération :	entre 700 € et 1300 € suivant la formation.
Compétences requises : - Vision par ordinateur, Traitement/Compression d'image - C++ (OpenCV, Qt), Python - Windows, linux/gcc, cmake, git - Utilisation d'un framework d'apprentissage profond (Tensorflow, PyTorch)	