CEA List
Laboratoire de Vision et Apprentissage pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contact Astrid SABOURIN
Guillaume LAPOUGE
Tél +33 (0)1 69 08 33 63
E-mail Astrid.sabourin@cea.fr
guillaume.lapouge@cea.fr

THESE 2022 Réf : LVA-22-T1

Fusion de caractéristiques pour la ré-identification multi-vues par apprentissage profond

Présentation du laboratoire d'accueil

Basé à Paris-Saclay, le CEA List est l'un des quatre instituts de recherche technologique de CEA Tech, direction de la recherche technologique du CEA. Dédié aux systèmes numériques intelligents, il contribue au développement de la compétitivité des entreprises par le développement et le transfert de technologies.

L'expertise et les compétences développées par les 800 ingénieurs-chercheurs et techniciens du CEA List permettent à l'Institut d'accompagner chaque année plus de 200 entreprises françaises et étrangères sur des projets de recherche appliquée s'appuyant sur 4 programmes et 9 plateformes technologiques. 21 start-ups ont été créées depuis 2003.

Labellisé Institut Carnot depuis 2006, le CEA List est aujourd'hui l'institut Carnot Technologies Numériques

Le Laboratoire de Vision et Apprentissage pour l'analyse de scène (LVA) mène ses recherches dans le domaine de la Vision par Ordinateur (Computer Vision) selon quatre axes principaux :

- La reconnaissance visuelle (détection et/ou segmentation d'objets, de personnes, de patterns; détection d'anomalies; caractérisation)
- L'analyse du comportement (reconnaissance de gestes, d'actions, d'activités, de comportements anormaux ou spécifiques pour des individus, un groupe, une foule)
- L'annotation intelligente (annotation à grande échelle de données visuelles 2D/3D de manière semi-automatique)
- La perception et la décision (processus de décision markovien, navigation)

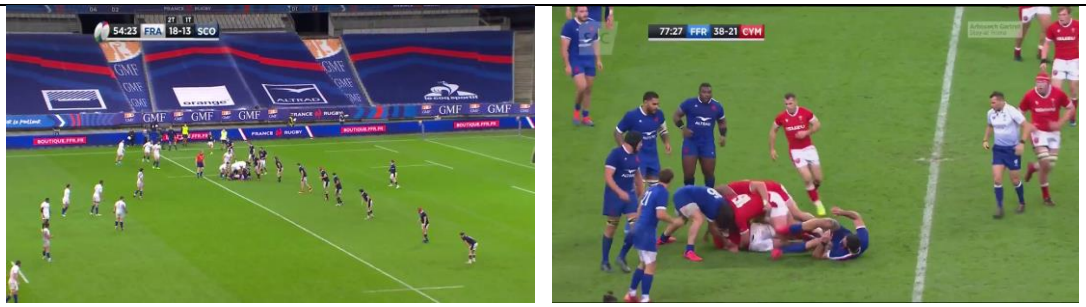
Problématique

De nombreuses applications de suivi telles que la vidéo protection de piétons, le monitoring de vaches dans une ferme ou encore l'analyse automatique de match de sports collectifs reposent sur la bonne ré-identification des individus dans un réseau de caméras. Cependant, il peut y avoir une grande disparité en termes de point de vue sur les individus en fonction de la caméra considérée.

Prenons l'exemple d'un match de rugby. Certaines caméras filment le match avec un plan large permettant de voir une grande partie du terrain et donc de localiser les individus facilement. Cependant la résolution des joueurs dans ces vues ne permet pas de les reconnaître aisément. Au contraire, d'autres caméras suivent l'action de façon beaucoup plus rapprochée. Il est plus facile d'identifier les joueurs dans cette vue. Un matching de ces deux types de vues permettrait de fusionner les informations de localisation à celles de ré-identification nécessitant plus de résolution. Cette association est loin d'être triviale du fait que les caméras sont en mouvement et qu'elles filment l'action avec des points de vue différents.

CEA List
Laboratoire de Vision et Apprentissage pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contact Astrid SABOURIN
Guillaume LAPOUGE
Tél +33 (0)1 69 08 33 63
E-mail Astrid.sabourin@cea.fr
guillaume.lapouge@cea.fr



Exemple d'une vue « plan large » à gauche et d'une vue « rapprochée » à droite

L'utilisation des réseaux de neurones permet de résoudre des tâches de vision de plus en plus complexes mais des limitations sont notables lorsqu'il s'agit d'apprendre des caractéristiques robustes aux changements de point de vue et aux changements d'échelle. Dans notre cas, extraire des caractéristiques de ré-identification robustes à l'incidence et au niveau de zoom sur les individus permettrait de :

1) reconnaître un même individu quel que soit le point de vue sur celui-ci ;
2) mettre en correspondance différentes caméras qui observent les mêmes individus de deux points de vue différents.

Remplir ces deux objectifs permettrait de robustifier le tracking multi-caméras des individus.

Objectifs

L'objectif de cette thèse est de proposer une méthode de fusion des caractéristiques multi-vues d'une même scène. L'approche finale devra tourner en temps réel en prenant en compte l'apparence visuelle des individus et leur position relative. La technologie développée pourra être testée sur des matchs de rugby dont les caméras sont synchronisées, pour du suivi individualisé de joueurs à long terme.

Pistes de recherche

Les travaux du doctorant pourront s'échelonner de la façon suivante :

- Dans un premier temps, le doctorant pourra proposer une amélioration des méthodes de ré-identification actuelles pour obtenir des caractéristiques 3D invariantes au changement de point de vue. Les travaux pourront s'inspirer de méthodes de reconnaissances d'action telle que [Piergiovanni21] qui utilise un modèle d'auto-calibration [Iyer18] capable de prédire à partir d'une image, les paramètres extrinsèques de la caméra. Ces paramètres peuvent être ensuite utilisés pendant l'apprentissage afin d'obtenir des caractéristiques 3D indépendants au point de vue. A l'inférence, le modèle d'auto calibration n'est plus utilisé mais les caractéristiques extraites par le réseau sont robustes au changement de point de vue.

  Catégoriser la position de l'individu dans la scène est également une tâche permettant au réseau de mieux appréhender les changements de point de vue. Notamment, [HE21] propose de fournir cette information directement en entrée du réseau.

  Dans cette première partie, le doctorant travaillera sur des imagettes d'individus dont les tailles d'observation sont semblables.

CEA List
Laboratoire de Vision et Apprentissage pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contact Astrid SABOURIN
Guillaume LAPOUGE
Tél +33 (0)1 69 08 33 63
E-mail Astrid.sabourin@cea.fr
guillaume.lapouge@cea.fr

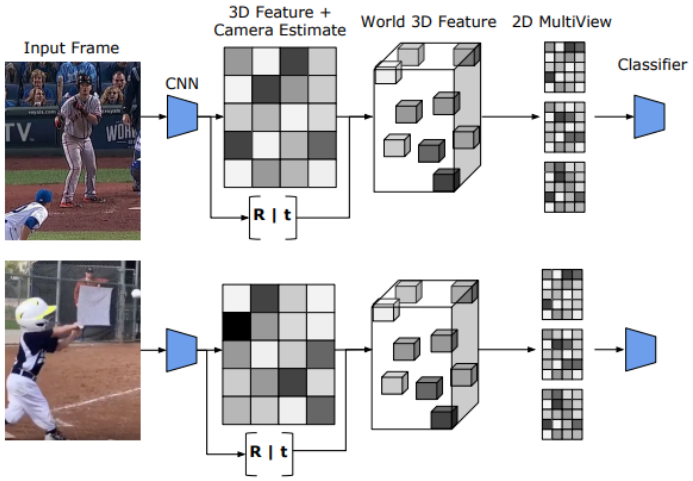


Figure issue de l'article [Piergiovanni21] : Exemple d'architecture permettant l'apprentissage de caractéristiques indépendantes au point de vue

- Dans un deuxième temps, le doctorant sera amené à prendre en compte les différences d'échelle tout en ayant à sa disposition l'intégralité de l'image. La difficulté sera d'associer les individus entre chaque vue via les caractéristiques de ré-identification 3D plus robustes apprises par la méthode proposée précédemment et en prenant en compte l'apparence globale de la scène comme la position des individus les uns par rapport aux autres.

- Enfin, la méthode développée pourra être intégrée temporellement à un algorithme de suivi multi-vue et testée dans le cadre du suivi long terme des joueurs d'un match de rugby. Pour ce faire, le doctorant pourra s'appuyer sur des solutions sur étagère de tracking [Zhang21] qui permettent de faire le suivi court terme, puis de déployer des techniques de re-identification robustes pour assurer le suivi-long terme.

Concernant les données, le doctorant aura à sa disposition des vidéos de matchs de rugby filmés par plusieurs caméras synchronisées, les datasets publics de sport tel que SoccerNet [Cioppa22] ou le premier dataset de suivi de joueurs de rugby à 7 [Maglo22] ainsi que les datasets publics de ré-identification pour la vidéo protection tel que Market1501 [Zheng15].

Références:

[Maglo22] Maglo, A., Orcesi, A., & Pham, Q. C. (2022). Efficient tracking of team sport players with few game-specific annotations. arXiv preprint arXiv:2204.04049.

[Fu22] Fu, D., Chen, D., Yang, H., Bao, J., Yuan, L., Zhang, L., ... & Chen, D. (2022). Large-Scale Pre-training for Person Re-identification with Noisy Labels. arXiv preprint arXiv:2203.16533.

[Lan22] Lan, L., Teng, X., Chi, H., & Zhang, X. (2022). Multi-scale Knowledge Distillation for Unsupervised Person Re-Identification. arXiv preprint arXiv:2204.09931.

[Zhang22] Zhang, X., Li, D., Wang, Z., Wang, J., Ding, E., Shi, J. Q., ... & Wang, J. (2022). Implicit Sample Extension for Unsupervised Person Re-Identification. arXiv preprint arXiv:2204.06892.

[Iyer18] Iyer, G., Ram, R. K., Murthy, J. K., & Krishna, K. M. (2018, October). CalibNet: Geometrically supervised extrinsic calibration using 3D spatial transformer networks. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1110-1117). IEEE.

CEA List
Laboratoire de Vision et Apprentissage pour l'analyse de scène
Centre de Saclay 91191 Gif-sur-Yvette France
http://www.kalisteo.eu

Contact Astrid SABOURIN
Guillaume LAPOUGE
Tél +33 (0)1 69 08 33 63
E-mail Astrid.sabourin@cea.fr
guillaume.lapouge@cea.fr

[Piergiovanni21] Piergiovanni, A. J., & Ryoo, M. S. (2021). Recognizing actions in videos from unseen viewpoints. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4124-4132).

[Cioppa22] Cioppa, A., Deliège, A., Giancola, S., Magera, F., Somers, V., Cheng, Z., ... & Van Droogenbroeck, M. (2022). Soccer Player Tracking, Re-Identification, Camera Calibration and Action Spotting-SoccerNet Challenge 2022.

[Zheng15] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable person re-identification: A benchmark. In Proceedings of the IEEE international conference on computer vision (pp. 1116-1124).

[Zhang21] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Yuan, Z., Luo, P., Liu, W., Wang, X. (2021). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. arXiv:2110.06864.

[He21] He, S., Luo, H., Wang, P., Wang, F., Li, H., Jiang, W., 2021. TransReID: Transformer-based Object Re-Identification. arXiv:2102.04378.

Niveau demandé :	Ingénieur, Master 2
Durée :	3 ans
Rémunération :	entre 1800 € et 2000 €.

Compétences requises :

- Vision par ordinateur
- Apprentissage automatique (deep learning)
- Reconnaissance de formes
- C/C++, Python
- La maîtrise d'un framework d'apprentissage profond (en particulier Tensorflow ou PyTorch) est un plus**.**