

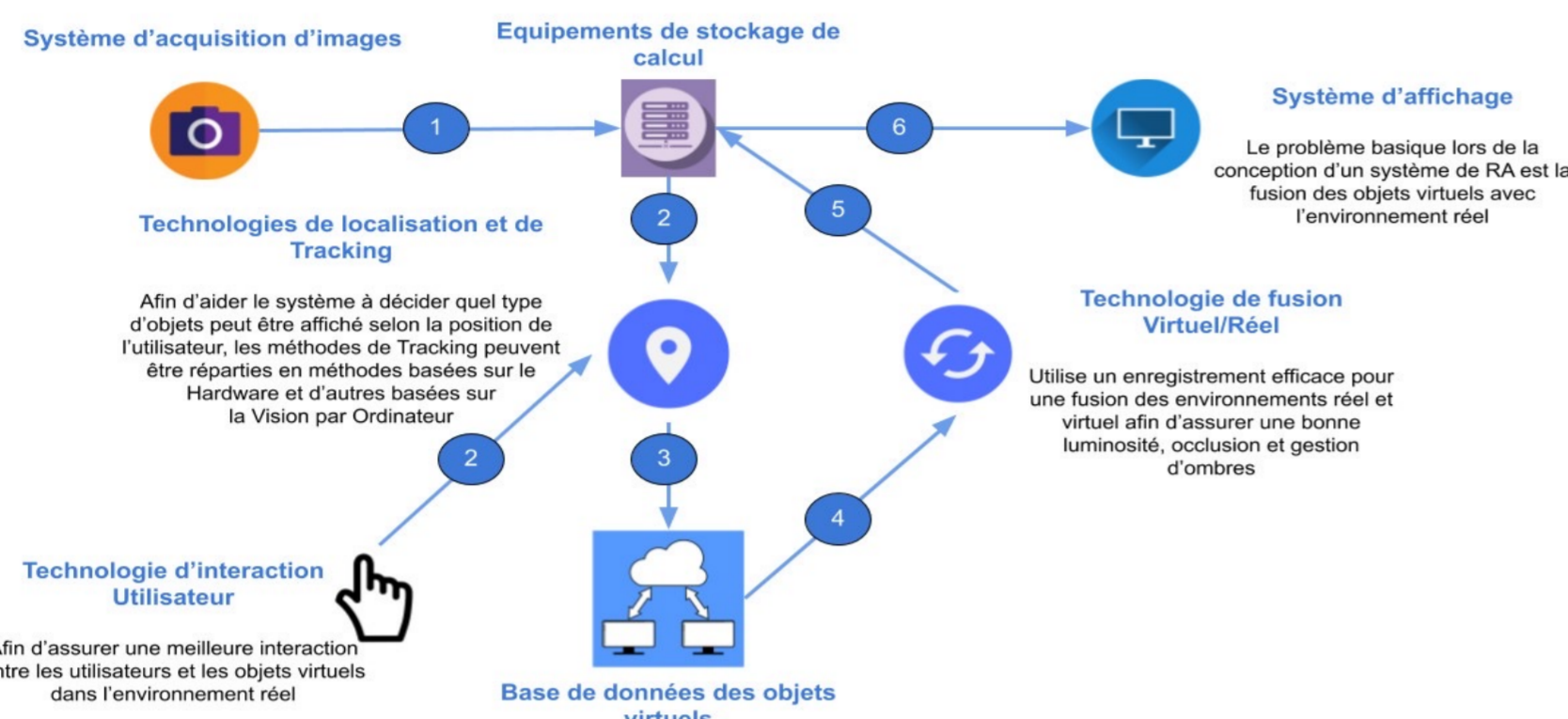
# Deep Learning pour le Tracking en Réalité Augmentée

**Nom et Prénom** : Samir Si-Mohammed  
**Spécialité** : Systèmes Informatiques  
**Organisme d'accueil** : LMCS-ESI (Alger)  
**Encadré par** : Mme Karima Benatchba (ESI)



## Réalité Augmentée

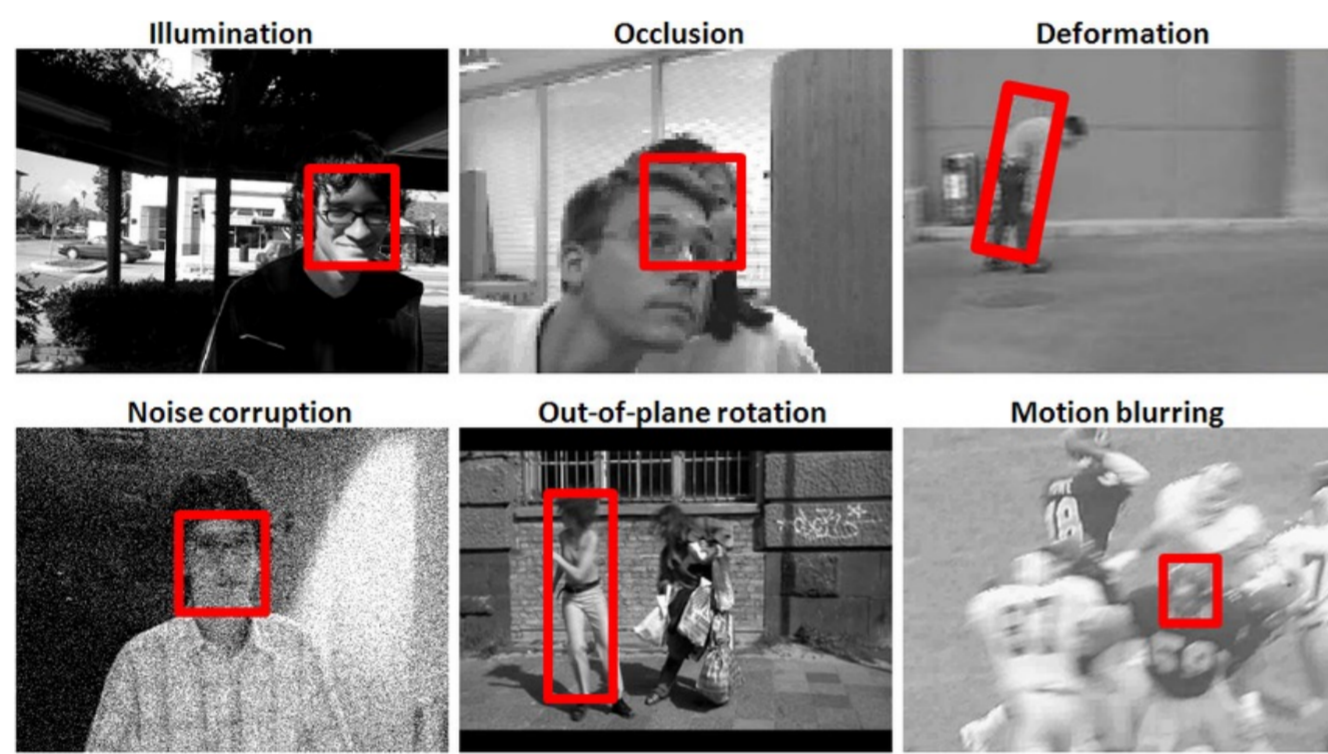
1. La combinaison des contenus virtuel et réel
2. L'interaction en temps réel
3. Intégration 3D des objets virtuels



Processus de RA (Cheng et al., 2018)

## Tracking en Réalité Augmentée

- Environnement dynamique
- Multitude d'objets
- Fusion des objets réels et virtuels difficile
- Immersion totale de l'utilisateur



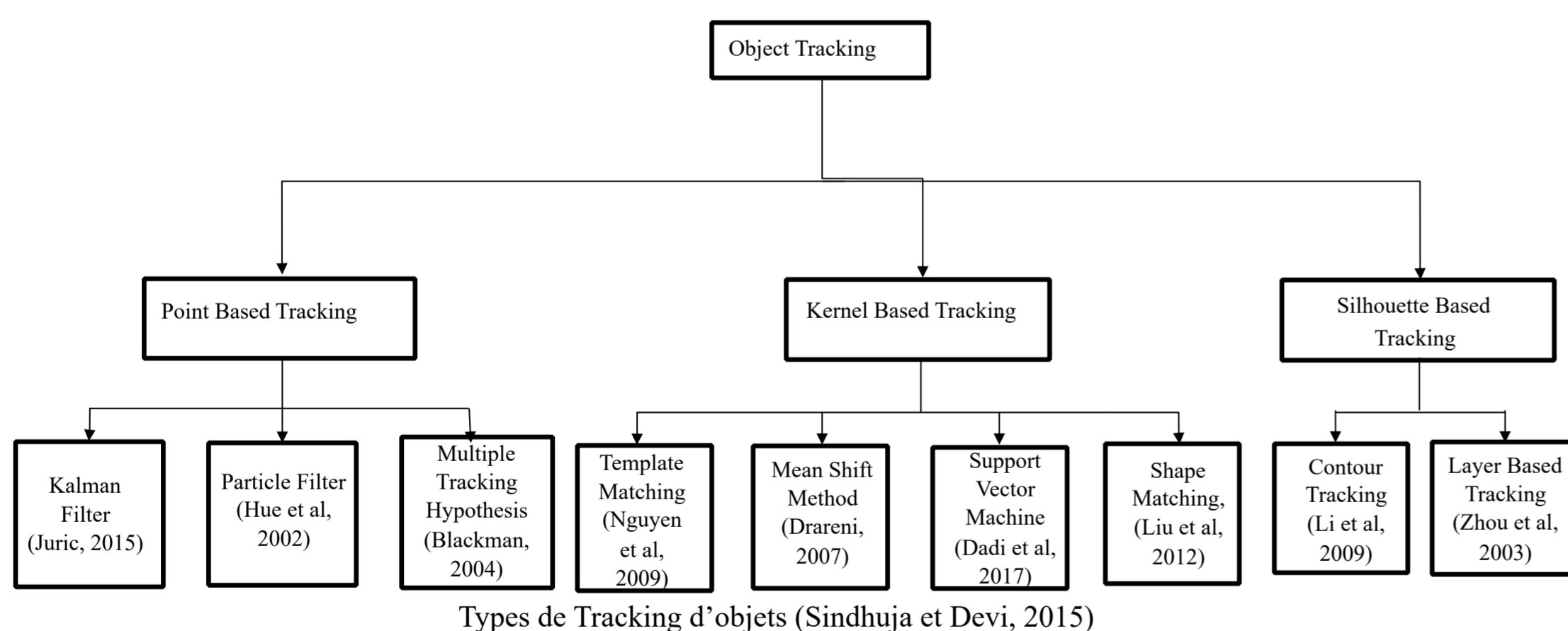
Difficultés du Tracking (Li et al., 2013)

### Problématique

- La RA est une technologie permettant à l'utilisateur de voir le monde réel, avec des objets virtuels superposés au-dessus de ceux du monde réel.
- Le tracking consiste à garder la trace d'un l'objet dans un environnement en détectant sa position, et ce pour permettre une meilleure intégration des objets virtuels dans l'environnement réel.

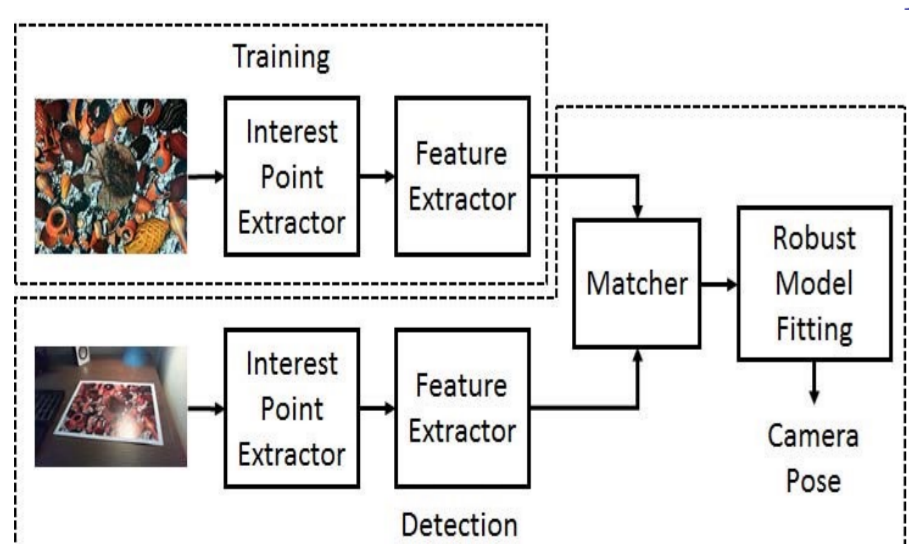
→ Comment effectuer le tracking en Réalité Augmentée de manière efficace afin d'augmenter l'immersion de l'utilisateur dans l'expérience ?

## Aperçu Général de l'état de l'art

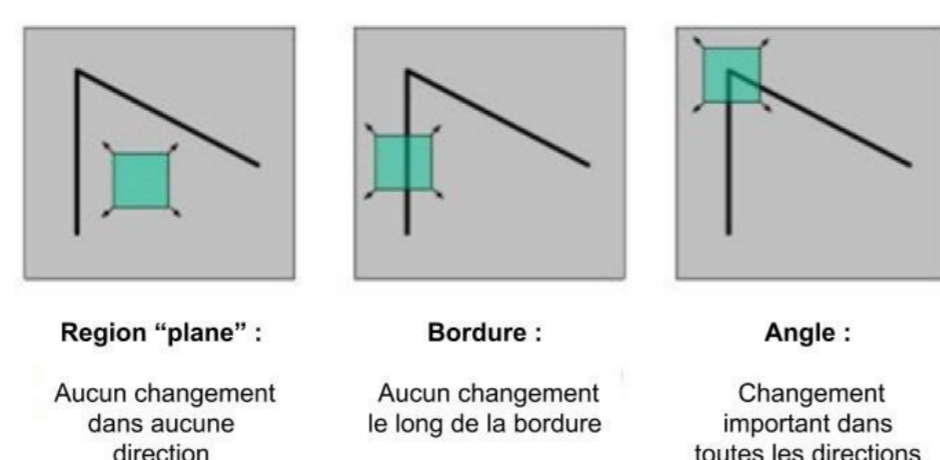


Types de Tracking d'objets (Sindhuja et Devi, 2015)

## Deep AR



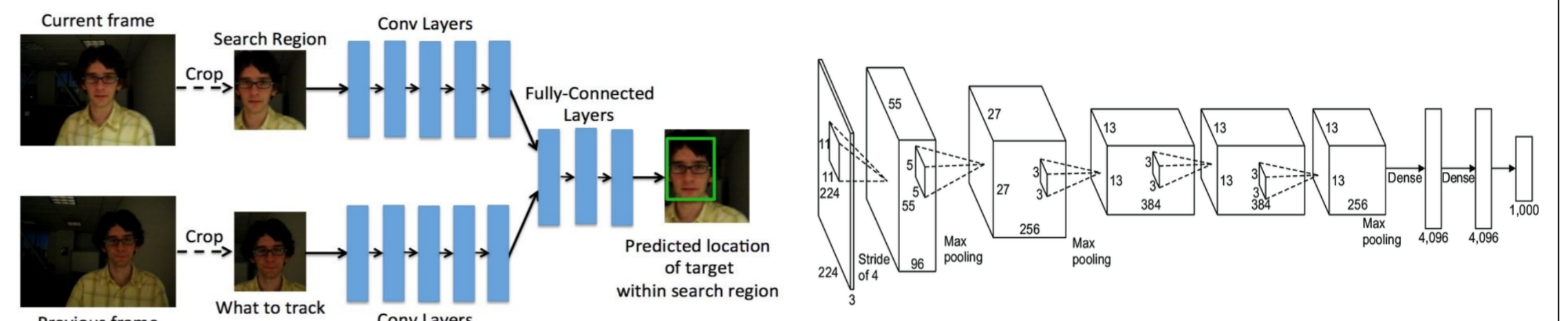
Etapas de détection en Tracking RA (Akgul et al., 2016)



Détecteur de Harris (Jieyang Hu, 2015)

- Le détecteur de Harris permet de détecter de manière très efficace les points d'angle dans une image. Les points d'angle sont considérés comme des points clé dans une image.
- Les images en entrée sont d'abord passées à travers un extracteur de points d'intérêt dans chaque image. Puis, celles-ci contenant les points d'intérêt sont ensuite passées à travers un extracteur de caractéristiques.

## GOTURN



Architecture CaffeNet (Hyung Lee, 2018)

Architecture GOTURN (LearnOpenCv, 2018)

- On cherche à prédire la position de l'objet au moment  $t$  à partir de la position directement précédente au moment  $t-1$
- L'entraînement se fait en soumettant au réseau deux images successives chacune ayant la cible correctement entourée. La sortie du réseau représente les coordonnées 2D du rectangle entourant la cible

## Quelques méthodes

- Lee, Y. H., & Lee, W. B. (2019, July). Object Detection and Tracking Based on Deep Learning. In International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (pp. 629-635). Springer, Cham.
- Rao, J., Qiao, Y., Ren, F., Wang, J., & Du, Q. (2017). A mobile outdoor augmented reality method combining deep learning object detection and spatial relationships for geovisualization. Sensors (Switzerland), 17(9). <https://doi.org/10.3390/s17091951>
- Brunetti, A., Buongiorno, D., Trotta, G. F., & Bevilacqua, V. (2018). Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing, 300, 17-33. <https://doi.org/10.1016/j.neucom.2018.01.092>

## Métriques d'évaluation

- Le Ground Truth désigne le fait d'extraire un ensemble d'informations objectives concernant le contenu d'une région quelconque, comme la localisation, la taille, la couleur...etc. afin de comparer celles-ci avec les informations générées par des images satellite par exemple.
- L'objectif des métriques est d'estimer le nombre de valeurs correctes prédites d'un algorithme par rapport aux valeurs correctes du Ground Truth. On peut estimer durant le Tracking les différentes valeurs suivantes, et les comparer au Ground Truth comme fait dans les méthodes de Machin Learning :
  - Segmentation : Déterminer la qualité de la forme finale de la cible prédite.
  - Performance de détection : Présence ou pas de l'objet cible dans l'image.
  - Complétude du Tracking : Evaluer la qualité du suivi de l'objet dans la vidéo.
  - Classification de la cible : Déterminer la classe à laquelle l'objet cible appartient.

## Perspectives

- La plupart des méthodes de tracking dans la littérature, utilisant le deep learning, ne sont pas destinées à la Réalité augmentée, mais au simple suivi d'objets. Toutefois, il serait intéressant qu'une fois ces méthodes établies, elles soient utilisées pour les systèmes RA, à condition de bien surveiller leur temps d'exécution ainsi que leur consommation d'énergie, qui sont deux facteurs importants pour le matériel RA.
- Dans certains cas ces images sont traitées comme si elles étaient indépendantes (comme pour DeepAR), c'est-à-dire que nous n'exploitons pas assez le fait que les images se suivent et qu'il y ait une relation de précedence entre elles. L'application des réseaux de neurones récurrents permettrait d'exploiter la succession des images, c'est-à-dire que la position d'un objet à l'instant  $t$  influera inéluctablement sur sa position à l'instant  $t+1$ . Certaines recherches ont d'ores et déjà été effectuées dans ce domaine, comme (Fang, 2016) qui propose l'utilisation de RNN pour la détection de personnes mobiles. Néanmoins sa solution présente quelques lacunes lorsque les personnes sont habillées de la même couleur par exemple, ou plus généralement ont la même apparence. Une autre étude est celle de (Milan et al., 2017) qui utilise une architecture assez connue des RNN qui est Long Short Term Memory (LSTM) qui est surtout utilisée pour la reconnaissance de caractères et de paroles. Le réseau développé présente toutefois des résultats considérés satisfaisants.

## Références

- Akgul, O., Penekli, H. I., & Genc, Y. (2017). Applying Deep Learning in Augmented Reality Tracking. Proceedings - 12th International Conference on Signal Image Technology and Internet-Based Systems, SITIS 2016, 47-54.
- Held, D., Thrun, S., & Savarese, S. (2016). GOTURN: Learning to Track at 100 FPS with Deep. Eccv 2016, 749-765. Retrieved from <http://davheld.github.io/GOTURN/GOTURN.html>
- Sindhuja, G., & M, R. D. S. (2015). A Survey on Detection and Tracking of Objects in Video Sequence. International Journal of Engineering
- Fang, K. (2016). Track-RNN: Joint Detection and Tracking Using Recurrent Neural Networks. 29th Conférence on Neural Information Processing Systems (NIPS 2016), (Nips). Retrieved from [https://web.stanford.edu/class/cs231a/prev\\_projects\\_2016/final\\_report\(7\).pdf](https://web.stanford.edu/class/cs231a/prev_projects_2016/final_report(7).pdf)
- Milan, A., Rezatofghi, S. H., Dick, A., Reid, I., & Schindler, K. (2017). Online multi-target tracking using recurrent neural networks. 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 4225-4232.