

**Sujet de stage Master recherche**  
**Utilisation de la sémantique pour le calcul de pose**  
**Equipe Magrit (LORIA/INRIA, Nancy)**  
**Marie-Odile Berger et Gilles Simon**

## 1 Contexte

La réalité augmentée vise à accroître la perception qu'un utilisateur a de son environnement en ajoutant dans son champ visuel des informations ou des objets graphiques. Le besoin d'afficher des éléments au bon endroit dans les images et de les intégrer de façon réaliste dans l'environnement nécessite de connaître la position de la caméra à chaque instant. Le calcul de pose a été longtemps basé uniquement sur la mise en correspondance entre points d'intérêt 2D extraits dans l'image et points 3D d'un modèle de la scène [3]. En pratique, la génération d'hypothèses est basée sur les similarités photométriques et le taux d'hypothèses incorrectes peut être assez élevé. Le calcul de la pose est donc fait par échantillonnage : 4 correspondances 2D/3D permettant le calcul de la pose, on génère aléatoirement des quadruplets de correspondances et on évalue la qualité de la solution trouvée. Traditionnellement, cette évaluation repose sur la taille de l'ensemble de consensus, c'est-à-dire l'ensemble des points en accord avec cette pose. En pratique, dans des contextes applicatifs complexes avec de nombreux motifs répétés, de forts changements de points de vue ou des changements d'illumination dans la scène, les hypothèses 2D/3D ne sont pas fiables et en petit nombre. La taille des ensembles de consensus est alors faible rendant l'identification de la solution optimale parmi les solutions possibles difficiles.

## 2 Sujet

Avec l'avènement des méthodes de reconnaissance d'objets, on dispose désormais de méthodes assez fiables d'étiquetage sémantique (segnet [1],...). Ces méthodes sont par ailleurs relativement invariantes aux changements d'illumination. Plusieurs tentatives ont donc été faites pour utiliser cette information sémantique pour le calcul de pose. Certaines l'utilisent comme modalité principale, le recalage modèle/image étant fait uniquement sur la sémantique [2]. D'autres l'utilisent comme modalité secondaire, la sémantique servant uniquement à évaluer la qualité de la solution [4].

Ce sujet vise à ouvrir une réflexion sur la meilleure façon d'utiliser la sémantique pour le calcul de pose. On travaillera dans le cadre de scènes modélisées par un ensemble de points (SFM). On considérera plus particulièrement le contexte urbain. Plusieurs étapes sont envisagées :

- Construction d'un modèle sémantique 3D à partir des images : En particulier, il sera nécessaire de raffiner un réseau CNN existant pour l'adapter aux primitives urbaines utilisées (fenêtres, portes,...).
- Proposition et évaluation d'une méthode pour évaluer une hypothèse de pose basée sur les informations sémantiques et étude de la pertinence de cette méthode. Cette étape pourra s'appuyer sur [4]
- Les informations sémantiques ne sont en pratique pas invariantes au changement d'échelle spatial. Le candidat devra proposer et mettre en oeuvre une méthode pour tenir compte de ce problème.

**Lieu du stage :** Nancy

**Pré-requis :** des connaissances en traitement d'image, vision par ordinateur et apprentissage seront appréciées.

Une maquette implantant les méthodes proposées sera réalisées en matlab.

## Références

- [1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet : A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

- [2] Antoine Fond. *Image-based localization in urban environment : application to augmented reality*. Theses, Université de Lorraine, April 2018.
- [3] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Epnp : An accurate  $o(n)$  solution to the pnp problem. *International Journal of Computer Vision*, 81(2) :155, Jul 2008.
- [4] Carl Toft, Erik Stenborg, Lars Hammarstrand, Lucas Brynte, Marc Pollefeys, Torsten Sattler, and Fredrik Kahl. Semantic match consistency for long-term visual localization. In *The European Conference on Computer Vision (ECCV)*, September 2018.