

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



*RAPPEL : le dossier doit être transmis par l'organisme ou l'établissement auquel est rattaché le directeur de thèse.*

## 1. INTITULE DE LA THESE

- Intitulé de l'axe et de la problématique IRT concernés :

Axe robotique/cobotique – Positionnement centimétrique dans l'industrie Indoor

- Intitulé de la thèse :

Combinaison de SLAM visuel et de reconnaissance d'objets 3D basée deep learning pour une localisation grande et petite échelle au sein de l'atelier

- Etablissement d'enseignement supérieur où sera inscrit le doctorant :

Ecole Centrale de Nantes

- Ecole doctorale :

SPI (nouveau nom de l'ED SPIGA depuis le 01/09/2017)

- Laboratoire où s'effectuera la thèse :

AAU UMR 1563 CNRS – ECN – ENSA Nantes – ENSA Grenoble

- Directeur du laboratoire :

Daniel SIRET

- Directeur de la thèse :

Guillaume MOREAU

- Collaboration avec d'autres laboratoires dans le cadre de la thèse (nature de la collaboration) :

- Nom des autres laboratoires impliqués dans la collaboration :

## 2. CANDIDAT ENVISAGE (joindre un CV)

- Nom :

- Prénom :

- Date de naissance :

- Nationalité :

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



## 3. CANDIDAT ENVISAGE N°2 (joindre un CV)

- Nom :
- Prénom :
- Date de naissance :
- Nationalité :

## 4. DESCRIPTION DU SUJET (1 page)

Un des problèmes majeurs pour le développement des technologies de Réalité Augmentée (RA) sur site en environnement industriel concerne les contraintes fortes des milieux industriels : conditions d'éclairage très variables et non contrôlables, présence de matériaux réfléchissants (métaux) et absence de couleurs ou de textures, environnement changeant (déplacement des matières premières sur le chantier), utilisation d'outils encombrants, potentiellement dangereux, etc.

Les enjeux des applications de RA en milieu industriel sont doubles : 1) être capable de localiser le dispositif de RA (tablette, casque, etc.) dans l'atelier (i.e. à grande échelle) de manière précise en 6D ; 2) être capable d'augmenter de manière précise (précision sous-centimétrique) les objets (i.e. à petite échelle) nécessaires aux opérateurs (outils, pièces à usiner, etc.).

Dans ce contexte difficile, les outils classiques de RA fonctionnent mal ou ne répondent qu'à une partie de ces enjeux. Les technologies basées SLAM visuel, en produisant une reconstruction 3D temps-réel de l'environnement dans lequel un dispositif évolue, permettent de se repérer à grande échelle dans des environnements industriels mais ne permettent pas de faire de la reconnaissance d'objets dans la scène ni de présenter de manière suffisamment précise les augmentations.

Au contraire, les techniques de l'état de l'art en RA permettant d'obtenir des augmentations avec des précisions très importantes (inférieures au centimètre), mais ne permettent pas de faire une localisation précise à grande échelle.

L'intérêt de cette thèse est donc de proposer une approche hybride où nous souhaitons permettre d'effectuer une localisation à grande échelle (basée sur des techniques de SLAM visuel) mais également d'être capable de pourvoir faire de la reconnaissance d'objets (outils, pièces, etc.) et d'obtenir leurs positions 3D et orientations 3D afin de pouvoir faire de la RA sur site. La partie reconnaissance d'objet se base sur une idée très récente couplant de l'apprentissage profond (*deep learning*) avec des méthodes de reconnaissance d'objets et qui semble prometteuse.

Le résultat visé est de proposer une solution permettant de guider les opérateurs à grande échelle en environnement extérieur de manière précise (partie SLAM visuel), puis d'être également capable de présenter des augmentation pertinentes pour des tâches requérant une précision fine (partie reconnaissance d'objets basée deep learning). Ces deux tâches ayant lieu dans des environnements inconnus et non contrôlés (en termes de conditions d'éclairage, de contenu etc.).

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



Un scénario d'application consisterait à être capable d'afficher un itinéraire en réalité augmentée à un opérateur sur un chantier de construction le dirigeant vers sa prochaine tâche à effectuer. Une fois arrivé là où il doit réaliser sa tâche, on souhaite alors être capable de le guider dans les opérations qu'il a à réaliser en lui affichant par exemple les prochaines étapes de sa tâche (choix d'outil, action à effectuer, etc.).

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



## 5. ETAT DE L'ART (2- 3 pages)

L'utilisation des méthodes de machine learning pour des problèmes de vision par ordinateur n'est pas fondamentalement nouvelle, en effet, c'est l'idée même sous-jacente à celle de voiture autonome apparue dans les années 1980. Toutefois, un regain notable d'intérêt concernant les méthodes de deep learning est apparu depuis quelques années, notamment car ces méthodes obtiennent des résultats meilleurs que ceux de l'état de l'art jusqu'à lors.

C'est en particulier vrai pour les problèmes de détection de points d'intérêt dans les images où les algorithmes qui utilisaient jusqu'à maintenant des descripteurs manuels, comme par exemple SIFT [Lowe2004], SURF [Bay2006] ou ORB [Rublee2011] sont maintenant surpassées par des méthodes de machine learning, en particulier utilisant des réseaux de neurones convolutionnels (Convolutional Neural networks – CNNs), voir par exemple [Verdie2015] ou [Simo-Serra2015] ou des combinaisons de plusieurs couches de CNNs, on parle alors de Deep Learning, voir [Yi2016].

Les technologies liées au deep learning tendent également à se démocratiser grâce à la mise à disposition de solutions logicielles « clé en main » comme TensorFlow<sup>1</sup> de Google ou PyTorch<sup>2</sup>.

Concernant la détection automatique d'objets dans des images, ou le calcul de pose d'une caméra, il existe une littérature très fournie, nous ne présenterons ici que les travaux majeurs et en particulier aux rares articles ayant attiré à l'utilisation de deep learning pour ce problème.

Il existe plusieurs manières d'effectuer de la reconnaissance d'objets, cela peut dépendre de la méthode utilisée bien entendu, mais également des données d'entrée du problème. En effet, certaines solutions peuvent être utilisées lorsque nous ne disposons uniquement d'images en couleur (image RGB) mais certaines requièrent une image RGB-D c'est à dire une image couleur avec information de profondeur (comme obtenue par exemple avec un dispositif de type Kinect).

Concernant les méthodes basées images uniquement, citons par exemple les méthodes basées points d'intérêt comme [Lowe2014] ou [Wagner2008]. A partir d'images RGB-D des méthodes se basent sur des votes à partir de paires de points 3D et de leurs normales [Drost2010], utilisent des arbres de décision [Lai2011], des appariements de patrons (template matching) [Wohlhart2015].

Dans ce domaine aussi, les méthodes liées aux CNNs tendent à se développer depuis 2015 à peu près. [Kendall2015] utilise un CNN pour prédire la pose 3D d'un objet à partir d'une image seule. [Gupta2015] utilise également un CNN mais à partir d'images RGB-D.

Plus récemment, [Rad2017] se base également sur un CNN et sur des images couleurs uniquement. Leurs résultats sont assez largement supérieurs à ceux des méthodes de l'état de l'art sur le même problème de calcul de pose 3D d'objets. Dans cette thèse nous comptons nous baser sur l'utilisation de CNNs pour la reconnaissance d'objets en environnements complexes. Gardons toutefois à l'esprit que les méthodes de deep learning, présentent des inconvénients non négligeables à savoir : le besoin d'avoir une très grande quantité de données et d'exemples.

---

<sup>1</sup> <https://www.tensorflow.org>

<sup>2</sup> <http://pytorch.org>

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



## Perception pour la localisation indoor

Pouvoir se localiser dans l'environnement complexe d'un atelier grâce à des capteurs et de l'intelligence embarquée constitue un processus important pour de nombreux systèmes qu'ils soient dédiés à la réalité augmentée ou à la robotique. Pour les applications de réalité augmentée, l'information de localisation consiste à identifier la position et/ou l'orientation du point de vue. Ces données sont importantes pour connaître la portion du monde perçue par l'utilisateur. Ceci permet essentiellement d'assurer par la suite une cohérence du mixage du réel et du virtuel. En effet, l'estimation des paramètres de localisation permet de d'aligner correctement les mondes réel et virtuel et ainsi de créer la vue augmentée.

Tout comme nous essayons de nous localiser avec des points d'intérêt persistants quand nous visitons une nouvelle ville (clocher d'église, place de la Mairie et son café du Midi, etc.), nous faisons en sorte qu'un système autonome naviguant adopte ce même comportement. Il détecte des zones caractéristiques avec ses capteurs, zones qui ont peu de chance d'évoluer au fil du temps et qu'il est capable de suivre ou de reconnaître dans les perceptions suivantes afin de pouvoir se localiser quelle que soit sa position.

SLAM est l'acronyme anglais pour Simultaneous Localization And Mapping, soit en français Localisation et Cartographie Simultanées. Au même instant, le système essaie de se localiser et de cartographier son environnement. Il se localise sur la base de la carte courante, et il complète cette carte de manière incrémentale. Il pourra mieux se repérer s'il passe par des lieux déjà visités. Un exemple est illustré dans la Figure 1. A chaque instant, le « système mobile » (équipé de capteurs pour percevoir l'environnement), perçoit des amers (notés L pour la formulation anglaise Landmark). S'il a déjà connaissance de ceux-ci, c'est-à-dire s'ils sont déjà dans la carte, il les observe à nouveau pour se localiser et pour affiner sa connaissance du monde, sinon il les initialise dans sa carte pour essayer de les réobserver aux itérations suivantes. Le principe est donc de localiser au mieux notre système à chaque instant.

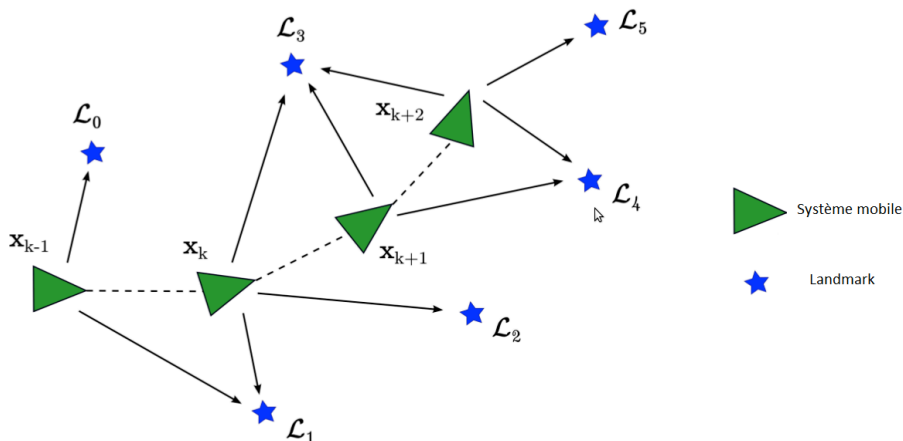


Figure 1 – Illustration du principe du SLAM : l'objectif est d'améliorer la connaissance sur la position de notre système. Pour cela, à chaque instant, il perçoit des percepts remarquables, ou amers, grâce à ses capteurs. Si ce sont des amers qu'il a déjà observés auparavant et donc qui font partie de sa carte, il effectue une ré-observation de ceux-ci. Ainsi, il affine simultanément sa connaissance sur sa position et sur celle de ces amers. Si ce sont des amers qu'il observe pour la première fois, il les intègre dans sa carte dans le but de les réobserver durant les itérations suivantes.

# Programme accélérateur de la recherche fondamentale sur le manufacturing

## Appel à proposition de sujets de thèse

### Dossier de candidature



Le SLAM est donc la méthode que nous allons utiliser pour localiser de notre système. Plusieurs types de capteurs tels que les caméras, lidars, etc., peuvent être utilisés pour le SLAM [Smith1986] [Lu1997]. Bien entendu, les caméras se sont rapidement déployées pour traiter la localisation d'un système mobile, et y devenir un capteur quasi-indispensable [Thrun2004] [Grisetti2007]. Les progrès technologiques (performances, compacité), la très large diffusion dans le grand public (coût) et le développement de nouveaux services sur les smartphones ont accéléré le processus. De plus, ces capteurs peuvent s'utiliser seuls, ou couplés pour former des systèmes multi-capteurs [Strasdat2010] [Kuemmerle2011]. Le choix pour réaliser un SLAM visuel multi-capteurs est donc vaste et se définit précisément par l'application concernée. Ici, nous souhaitons réaliser un SLAM qui soit robuste, ce qui impose l'utilisation d'un système multi-capteur.

Le problème du SLAM visuel se réfère souvent à celui de Structure From Motion (SFM) dont l'étude remonte aux années 80 et qui était un thème majeur de la Vision par Ordinateur. A cette époque, il était question de retrouver la géométrie d'une scène via une ou plusieurs caméras la percevant de différents points de vue de manière à pouvoir la reconstruire [Harris1988]. Ces études ont permis de comprendre le processus de vision et de poser les fondements de la géométrie visuelle [Hartley2000]. Elles n'abordaient que très rarement la problématique du SLAM. Même si SLAM et SFM traitent de thématiques proches, la vision dans la problématique du SLAM fut approfondie en 1997 par [Neira1997] s'appuyant sur les travaux de Chatila [Chatila1985], Smith [Smith1988] et Moutarlier [Moutarlier1989].

En 1998, Davison [Davison1998] propose l'utilisation d'un système de stéréovision couplé à des odomètres pour réaliser une cartographie tridimensionnelle de l'environnement. Percevoir un amer de deux points de vue différents au même instant permet de connaître plus précisément et plus rapidement la donnée de profondeur manquante dans l'approche mono-caméra. Cette approche sera étendue à un système stéréoscopique seul par Jung [Jung2003]. En 2003 le SLAM monoculaire connaît une grande avancée grâce aux travaux de Davison [Davison, 2003] qui propose un algorithme de SLAM, le MonoSLAM, utilisant une simple caméra et fonctionnant en temps réel sur un ordinateur grand public. En 2007 Klein et Murray [Klein2007], implémentent l'algorithme PTAM (Parallel Tracking and Mapping). Ils séparent le suivi de la caméra et la cartographie de l'environnement en deux processus parallèles (le mapping est dissocié du tracking et effectué en parallèle sur un thread différent).

Les approches multi-capteurs se généralisent en 2006, Kaess et Dellaert [Kaess2006], implémentent un algorithme du SLAM sur un système mobile équipé d'une ceinture de 8 caméras. En 2007, Mei et al. [Mei2007], utilisent une caméra omnidirectionnelle catadioptrique associée à un télémètre laser. Parmi les développements au niveau logiciel en matière de SLAM multi-capteur, l'algorithme RT-SLAM [Roussillon2011], permet de fusionner au sein d'une même application SLAM une ou plusieurs caméras, odomètres, centrales inertielles et GPS. Notons que de nombreux travaux traitent de perception multi-capteur en exploitant aussi des données GPS et des cartes de l'environnement [Bresson2013].

Trois problématiques complémentaires des techniques connues de SLAM, pourront être considérées :

(1) Le système pourra disposer d'un modèle a priori de l'environnement dans lequel il va évoluer; souvent ce modèle est donné par CAO. Il conviendra de recalibrer le modèle SLAM construit par perception avec ces connaissances a priori. On parle de SLAM contraint [Lee2007].

(2) Des cartes d'amers sous la forme d'objets suivis ou appariés dans les données sensorielles pourront être considérées. On parle de SLAM sémantique; une annotation par l'opérateur est d'abord nécessaire

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



pour associer un symbole, une utilité (ou affordance) aux différents objets ou lieux [Nuchter2008] [Rusu2009] [Civera2011]; cette annotation sera d'autant plus facile que le niveau de représentations est élevé (exploitation de connaissance a priori pour générer des modèles basés objets). Puis le système appliquera des techniques de reconnaissance pour identifier et localiser des objets.

(3) Enfin l'environnement dans lequel le système de RA va se déplacer est dynamique et évolutif; il pourrait contenir d'une part des objets mobiles non coopératifs (humains) et d'autre part des objets déplaçables (chaises, escabeau...). Nous devons filtrer les objets mobiles pour garantir une bonne localisation en utilisant par exemple des approches de type SLAMMOT (SLAM, plus Mobile Object Tracking) [Wang2004] [Marquez2012] [Lefaudeux2013].

## References:

[Bay2006] Bay H., Tuytelaars T., Van Gool L. 2006. « SURF: Speeded Up Robust Features ». In: Leonardis A., Bischof H., Pinz A. (eds) Computer Vision – ECCV 2006. ECCV 2006. Lecture Notes in Computer Science, vol 3951. Springer, Berlin, Heidelberg

[Bresson2013] G. Bresson, R. Aufrère and R. Chapuis. Consistent Multi-robot Decentralized SLAM with Unknown Initial Positions. Proc. FUSION, 2013.

[Chatila1985] R. Chatila, J. P.Laumond. Position referencing and consistent world modeling for mobile robots. Proc. ICRA1985.

[Civera2011] J. Civera, D. Galvez-Lopez, L. Riazuelo, J.D. Tardos and J.M.M. Montiel. Towards Semantic SLAM using a Monocular Camera. Proc. IROS2011.

[Davison1998] A. J. Davison and D. W. Murray. Mobile Robot Localisation Using Active Vision. ECCV, 1998.

[Davison2003] A. Davison. Real-time simultaneous localisation and mapping with a single camera. Proc. ICCV2003.

[Drost2010] Drost, Bertram, Ulrich, Markus, Navab, Nassir and Ilic, Slobodan. 2010. « Model globally, match locally: Efficient and robust 3D object recognition » In CVPR, 2010.

[Grisetti2007] G.Grisetti, C. Stachniss, W. Burgard. Improved Techniques for Grid Mapping with Rao-Blackwellized Particle Filters. IEEE Trans. Robotics, Vol.23, 2007.

[Gupta2015] S. Gupta, P. Arbeláez, R. Girshick and J. Malik. 2015. « Aligning 3D models to RGB-D images of cluttered scenes ». In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 4731-4740.

[Harris1988] C. Harris and M. Stephens. A combined corner and edge detector. In Proc. of Fourth Alvey Vision Conference, 1988.

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



- [Hartley2000] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press New York, NY, USA, 2000.
- [Jung2003] I.K. Jung and S. Lacroix. High Resolution Terrain Mapping using Low Altitude Aerial Stereo Imagery. ICCV, 2003.
- [Kaess2006] M. Kaess and F. Dellaert. Visual SLAM with a Multi-Camera Rig. Georgia Institute of Technology, 2006.
- [Kendall2015] Alex Kendall, Matthew Grimes and Roberto Cipolla. 2015. « PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization ». In Proceedings of the International Conference on Computer Vision (ICCV), 2015.
- [Klein2007] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. Proc. ISMAR2007.
- [Kuemmerle2011] R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, W. Burgard. g2o: A general framework for graph optimization. Proc. ICRA2011.
- [Lai2011] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. 2011. « A scalable tree-based approach for joint object and pose recognition ». In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI'11).
- [Lee2007] K. W. Lee, S. Wijesoma, J. Ibañez Guzmán. A constrained SLAM approach to robust and accurate localisation of autonomous ground vehicles. Robotics and Autonomous Systems, Vol. 55, 2007.
- [Lefaudeux2013] B. Lefaudeux. Détection, localisation et suivi des obstacles et objets mobiles à partir d'une plateforme de stéréovision. PhD, Ecole des Mines de Paris, 2013.
- [Lowe2004] David G. Lowe, « Distinctive image features from scale-invariant keypoints » in *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.
- [Lu1997] F.Lu, E.Milios. Globally Consistent Range Scan Alignment for Environment Mapping. Autonomous Robots, Vol.4, 1997
- [Marquez2012] D. Marquez-Gamez, M. Devy. Active visual-based detection and tracking of moving objects from clustering and classification methods. Proc. Advanced Concepts for Intelligent Vision Systems (ACIVS), 2012.
- [Mei2007] C. Mei and P. Rives Calibration between a Central Catadioptric Camera and a Laser Range Finder for Robotic Applications. ICRA 2006.
- [Moutarlier1989]. P. Moutarlier and R. Chatila. Stochastic multisensory data fusion for mobile robot location and environment modelling. ISRR, 1989.



# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



- [Neira1997] J. Neira, M.I. Ribeiro, and J.D. Tardos. Mobile Robot Localization and Map Building using Monocular Vision. 5th Int. Symp. on Intelligent Robotic Systems ' 97.
- [Nuchter2008] A. Nuchter and J. Hertzberg. Towards semantic maps for mobile robots. *Robotics and Autonomous Systems*, 56(11), November 2008.
- [Rad2017] Mahdi Rad & Vincent Lepetit. 2017. « BB8: A Scalable, Accurate, Robust to Partial Occlusion Method for Predicting the 3D Poses of Challenging Objects without Using Depth » in *Proceedings of the 2017 International Conference on Computer Vision (ICCV '17)*.
- [Roussillon2011] C.Roussillon, A.Gonzalez, J.Sola, J.M.Codol, N.Mansard, S.Lacroix, M.Devy. RT-SLAM: A generic and real-time visual SLAM implementation. *Lecture Notes in Computer Science*, Springer, 2011.
- [Ruble2011] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. « ORB: An efficient alternative to SIFT or SURF » In *Proceedings of the 2011 International Conference on Computer Vision (ICCV '11)*.
- [Rusu2009] R.B. Rusu, Z.C. Marton, N. Blodow, A. Holzbach, and M. Beetz. Model-based and learned semantic object labeling in 3D point cloud maps of kitchen environments. *Proc. IROS2009*.
- [Simo-Serra2015] E. Simo-Serra(\*), E. Trulls(\*), L. Ferraz, I. Kokkinos, P. Fua and F. Moreno-Noguer (\*: equal contribution). 2015. « Discriminative Learning of Deep Convolutional Feature Point Descriptors » In *International Conference on Computer Vision (ICCV)*, Santiago, 2015.
- [Smith1986] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. *Proc. Uncertainty in Artificial Intelligence (UAI)*, 1986.
- [Smith1988] R. Smith, M. Self and P. Cheeseman. A stochastic map for uncertain spatial relationships. *ISRR*, 1988.
- [Strasdat2010] H. Strasdat, J. Montiel, and A. Davison. Real-time monocular slam: Why filter? *Proc. ICRA2010*.
- [Thrun2004] S. Thrun. SLAM: a Survey. *Int. Journal on Robotics Research*, 2004.
- [Verdie2015] Y. Verdie, K. M. Yi, P. Fua and V. Lepetit. 2015. « TILDE: A Temporally Invariant Learned Detector » In *Computer Vision and Pattern Recognition (CVPR) 2015*
- [Wagner2008] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. 2008. « Pose tracking from natural features on mobile phones ». In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '08)*.
- [Wang2004] C.C. Wang. Simultaneous localization, mapping and moving object tracking. Ph.D. dissertation, Carnegie Mellon University, 2004.

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



[Wohlhart2015] Paul Wohlhart and Vincent Lepetit. 2015. « Learning Descriptors for Object Recognition and 3D Pose Estimation » In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2015

[Yi2016] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, Pascal Fua. 2016. « LIFT: Learned Invariant Feature Transform » in *Proceedings of the European Conference on Computer Vision (ECCV), 2016*

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



## 6. DESCRIPTION DES VEROUS SCIENTIFIQUES (2 pages)

Le problème dit de « tracking » en réalité augmentée a beau être connu depuis les débuts de ces technologies, il n'en reste pas moins un problème ouvert qui doit faire face à des nombreux challenges difficiles à résoudre séparément et encore plus quand ils sont combinés. Nous nous plaçons ici dans le cadre d'un opérateur dans un milieu industriel ayant besoin d'opérer dans un contexte de réalité augmentée, c'est-à-dire d'afficher dans un dispositif ad hoc des informations numériques *colocalisées* avec l'environnement réel : ainsi s'il regarde un endroit particulier de l'espace, disons une poutre, il doit voir s'afficher en surimpression des informations de nature géométrique (où trouer la poutre par exemple) ou sémantique (une étiquette reliée à la poutre indiquant son identifiant par exemple).

L'idée de cette thèse est également de non seulement être capable de fournir des informations géométriques ou sémantiques aux opérateurs mais également d'être capables de les aider dans la réalisation de leur tâche et plus précisément de les aider dans les tâches ne demandant que peu d'expertise. L'idée est ainsi de permettre aux opérateurs de se concentrer sur les tâches pour lesquelles ils présentent une réelle valeur ajoutée (p. ex. souder).

A cette fin, **le dispositif d'affichage doit déterminer avec précision, robustesse et rapidité la transformation rigide (position + rotation, aussi appelée pose de la caméra) qui le relie au monde réel.** Pour cela, il embarque un certain nombre de capteurs qui doivent répondre au problème précédent en y ajoutant des contraintes fortes en termes de **volume, de poids et d'autonomie électrique** et par conséquent de **puissance de calcul**. Le facteur **coût** est également à prendre en compte.

Parmi ces capteurs, on trouve notamment des GPS qui fournissent des informations de position d'une précision décimétrique à une fréquence d'environ 1Hz et encore uniquement en milieu extérieur, des accéléromètres qui mesurent des rotations qu'il faut donc intégrer 2 fois pour obtenir des positions/orientations avec des risques importants de dérive temporelle, des caméras optiques qui peuvent dans certains cas fournir des informations de position et d'orientation.

Le problème du calcul de pose de caméra pour la réalité augmentée est classiquement résolu en plaçant dans le champ de vision de la caméra un ensemble d'éléments connus et qui permettent ainsi de calculer la transformation entre la caméra et le monde. La plupart du temps ces éléments connus, aussi appelés marqueurs, sont des images planaires auxquelles on applique un prétraitement permettant de les reconnaître en temps réel dans le flux vidéo de la caméra. Il est également possible de se baser sur des objets 3D eux aussi connus à l'avance pour effectuer ce calcul de pose.

Ces solutions basées marqueurs fonctionnent de manière très efficace sont invisibles dans notre cas d'application, car difficilement applicables aux environnements extérieurs et inconnus (ou en constante évolution comme peut l'être un chantier de construction).

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



Comme nous l'avons présenté plus haut, l'idée pour la localisation grande échelle des opérateurs dans un environnement non contraint est d'utiliser des méthodes de SLAM. Malgré les avancées remarquables qu'a connu le SLAM dans les dernières années, faire du SLAM de manière robuste dans un environnement non connu à l'avance (ou en évolution rapide), en extérieur et peu contraint reste extrêmement difficile à faire. De nombreux algorithmes de l'état de l'art fonctionnent encore uniquement dans des conditions de laboratoire et perdent le tracking au bout de quelques secondes ou minutes.

Pour ce qui est de la reconnaissance ou détection d'objets, là aussi les méthodes de l'état de l'art ne sont pas capables de résoudre ces problèmes dans toutes les conditions, en particulier pour des conditions d'éclairage médiocres ou quand l'environnement contient énormément d'objets.

Pris indépendamment, les deux aspects centraux de la thèse (localisation grande échelle en environnement peu contraint) et reconnaissance d'objet en temps réel dans une scène complexe restent encore des problèmes ouverts. L'objectif final de cette thèse serait de proposer une combinaison (en s'accordant toutefois des hypothèses simplificatrices, au moins au début) ce qui présente un vrai challenge.

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



## 7. DESCRIPTION DU PROGRAMME DE TRAVAIL (1 page)

Le programme de travail proposé est le suivant :

- État de l'art sur les techniques de SLAM et de deep learning + montée en compétences sur ces eux domaines : 6 mois
- Proposition d'une solution technique de SLAM visuel pour la localisation en environnement extérieur : 2 mois
- Tests et validation de la solution SLAM proposée : 6 mois
- Étude et proposition d'un réseau de neurone pour la reconnaissance d'objets connus + génération éventuelle de jeu de données pour guider l'apprentissage supervisé : 6 mois
- Hybridation entre SLAM et Reconnaissance d'objets : proposition d'heuristiques et de tests : 4 mois
  - o quand utiliser le SLAM seul,
  - o quand lancer la reconnaissance d'objets
  - o etc.
- Choix d'un cas d'utilisation industriel : 1 mois (discussion avec partenaire industriel IRT à identifier)
- Proposition, implémentation et validation d'un démonstrateur : 8 mois
- Rédaction de thèse et préparation soutenance : 3 mois

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



## 8. RENSEIGNEMENTS SUR LE LABORATOIRE D'ACCUEIL (~1 page : thématiques, organisation, personnels)

*Le laboratoire AAU est une Unité Mixte de Recherche du CNRS associant les Écoles Nationales Supérieures d'Architecture de Grenoble et de Nantes et l'École Centrale de Nantes.*

Le laboratoire AAU est une Unité Mixte de Recherche du CNRS associant les Écoles Nationales Supérieures d'Architecture de Grenoble et de Nantes et l'École Centrale de Nantes.

Le laboratoire compte environ 110 membres répartis dans les deux équipes (Nantes et Grenoble), dont environ 1/3 de doctorants.

Depuis sa création en 1998, le laboratoire se développe de manière fortement interdisciplinaire, tant par sa composition (architectes, sociologues, informaticiens, anthropologues, urbanistes, géographes, physiciens, historiens, philosophes), que par les problématiques et enjeux auxquels il répond (architecture, environnement, ville), et par les outils qu'il conçoit et met en œuvre (méthodologies d'enquêtes in situ, modélisation et simulation des phénomènes d'ambiances, réalité virtuelle, etc.). Il dispose d'une équipe d'informaticiens avec une expérience reconnue dans les domaines de la réalité virtuelle et de la réalité augmentée.

Au sein du laboratoire AAU, l'équipe d'encadrement de la thèse est composée de deux informaticiens (Guillaume Moreau, PU et Jean-Marie Normand, MCF) de l'École Centrale de Nantes spécialistes en Réalité Virtuelle et Réalité Augmentée qui sont également membres associés de l'équipe INRIA Hybrid.

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



## 9. INSCRIPTION DE LA THESE AU SEIN DES AXES DE RECHERCHE DU LABORATOIRE (~20 lignes)

Au sein du laboratoire AAU, les thématiques de Réalité Virtuelle et Réalité Augmentée sont très importantes, en particulier le développement de techniques permettant de faire de la RA sur site est particulièrement intéressant dans l'optique de proposer des applications destinées aux architectes et urbanistes (visualisation sur site de bâtiments à construire, etc.). Au delà de ce premier domaine d'application, l'équipe a également travaillé avec plusieurs industriels en matière de RV/RA : Airbus et ses sous-traitants, DCNS (devenu récemment Naval Group), Artefacto, Ripple Motion... Les chercheurs impliqués dans cet axe réalité virtuelle/réalité augmentée sont associés à l'équipe INRIA Hybrid, ce qui leur permet à la fois de développer leurs compétences en matière de perception des environnements virtuels / augmentés et d'ouvrir vers de nouveaux domaines d'application.

## 10. LISTE DES THESES RECENTES OU EN COURS EN LIEN AVEC CE SUJET

L. YANG, "Recalage robuste à base de motifs de points pseudo aléatoires pour la réalité augmentée", Décembre 2016, Ecole Centrale de Nantes, direction : Guillaume Moreau, encadrement : Jean-Marie Normand, bourse MESR.

N. ANTIGNY, "Caractérisation du niveau d'incohérence de données géoréférencées du point de vue d'un piéton par mises en correspondance multicritères" soutenance prévue fin 2018 direction : Valérie Renaudin (IFSTTAR), encadrement : Myriam Servières (ECN), bourse ECN-IFSTTAR.

## 11. LISTE DES TRAVAUX LES PLUS SIGNIFICATIFS DU LABORATOIRE (PUBLICATIONS, OUVRAGES, COLLOQUES...) DANS LE CHAMP THEMATIQUE DE LA THESE

[Yang, Normand, Moreau 2015] L. Yang, J.-M. Normand, G. Moreau. "Local Geometric Consensus: a general purpose point pattern-based tracking algorithm", IEEE Transactions on Visualization and Computer Graphics, vol. 21, no. 11, pp. 1299-1308, Nov. 15 2015.  
doi: 10.1109/TVCG.2015.2459897

[Yang et al. 2016] L. Yang, H. Uchiyama, J.-M. Normand, G. Moreau, H. Nagahara and R. Taniguchi, "Real-time surface of revolution reconstruction on dense SLAM". In Proceedings of the 4th International Conference on 3D Vision (3DV) 2016, Stanford, USA, October 25-28, 2016.

[Yang, Normand, Moreau 2016] L. Yang, J.-M. Normand, G. Moreau, "Practical and precise projector-camera calibration". In Proceedings of the 15th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Merida, Mexico, September 19-23, 2016.

[Yang, Normand, Moreau 2015] L. Yang, J.-M. Normand and G. Moreau, "Augmenting off-the-shelf paper maps using intersection detection and geographical information systems," in Proceedings of the 14th IAPR International Conference on Machine Vision Applications (MVA 2015), Tokyo, 2015, pp. 190-193. doi: 10.1109/MVA.2015.7153164

# Programme accélérateur de la recherche fondamentale sur le manufacturing Appel à proposition de sujets de thèse Dossier de candidature



## COORDONNEES DU CONTACT AU SEIN DE L'ORGANISME OU DE L'ETABLISSEMENT

- Nom : *Guillaume MOREAU*

- Adresse : *1 rue de la Noe 44321 Nantes Cedex*

- Tel : *02 40 37 68 47*

- Adresse e-mail : [guillaume.moreau@ec-nantes.fr](mailto:guillaume.moreau@ec-nantes.fr)