

CAMÉRA 3D PAR DEPTH FROM DEFOCUS ET MACHINE LEARNING

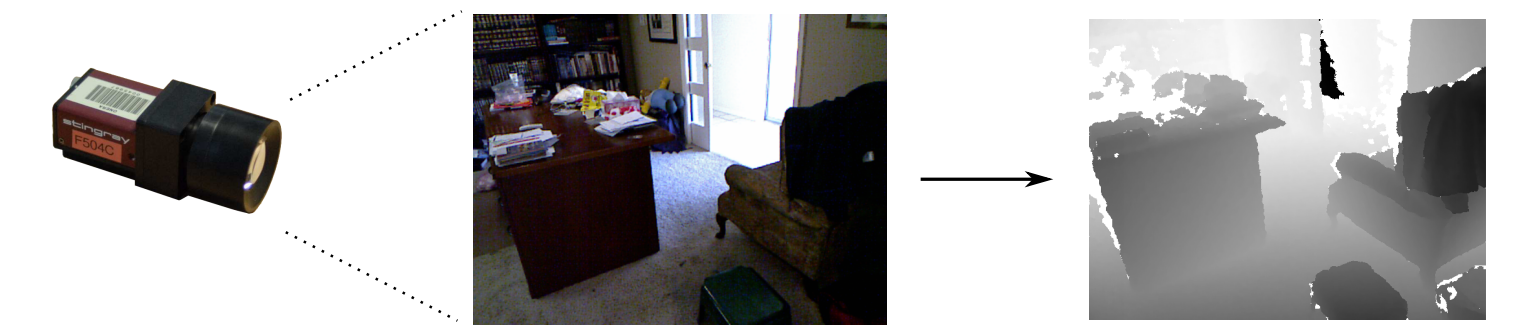
Marcela Pinheiro de Carvalho^{1,2}, Bertrand Le Saux¹, Pauline Trouvé-Peloux¹, Andrés Almansa², Frédéric Champagnat¹

ONERA/DTIS¹, Télécom ParisTech²



Contexte

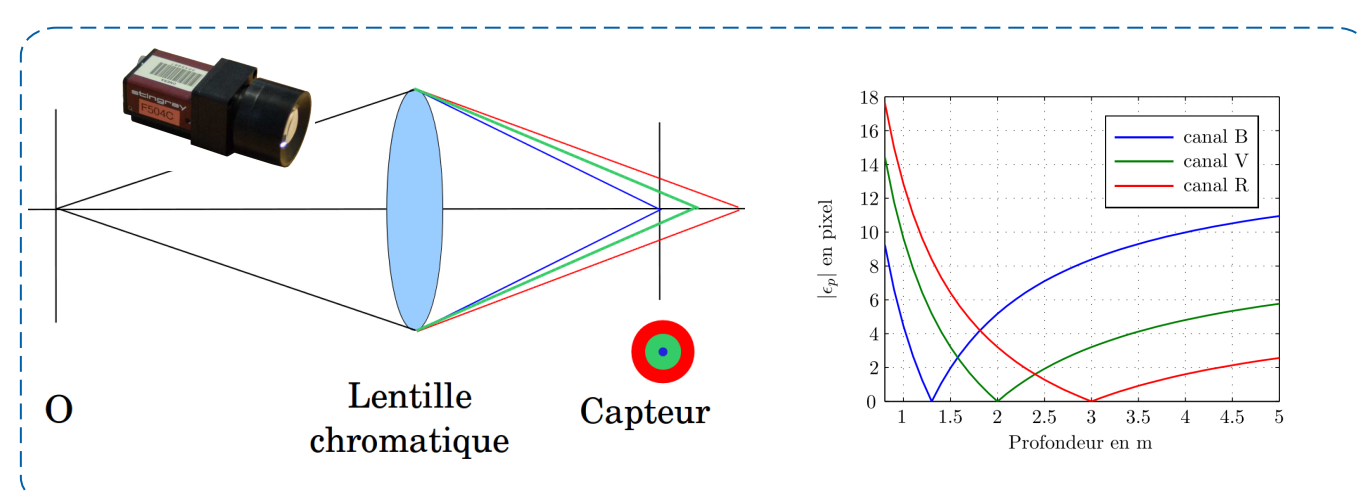
La **perception 3D** à partir d'une **seule image** est un problème important de la vision par ordinateur. Des approches actuelles pour l'estimation de profondeur demandent plusieurs capteurs et ont souvent des limitations selon l'environnement (soleil, texture). Le *machine learning* se présente actuellement avec des résultats étonnants sur plusieurs applications en vision par ordinateur. Nous proposons alors l'exploration des méthodes d'**intelligence artificielle** où la vision 3D peut possiblement atteindre des objectifs de **rapidité**, de **compacité** et de **coût**.



La thèse vise à concevoir un **système de vision 3D compact et rapide** pour des applications en **robotique**. Pour cela nous proposons de combiner un capteur dédié au *Depth from Defocus* à un algorithme d'estimation de profondeur issu du *Deep Learning*.

Depth from Defocus

Permet d'avoir une information 3D à partir d'un seul capteur passif à l'aide du **lien** entre le **flou de défocalisation** et la **position 3D**.



L'ONERA a développé CAM3D [1], dont l'optique est chromatique pour avoir des flous de défocalisation différents sur les canaux RVB.



Avantages

- Faible encombrement.
- Faible consommation.
- Faible coût.

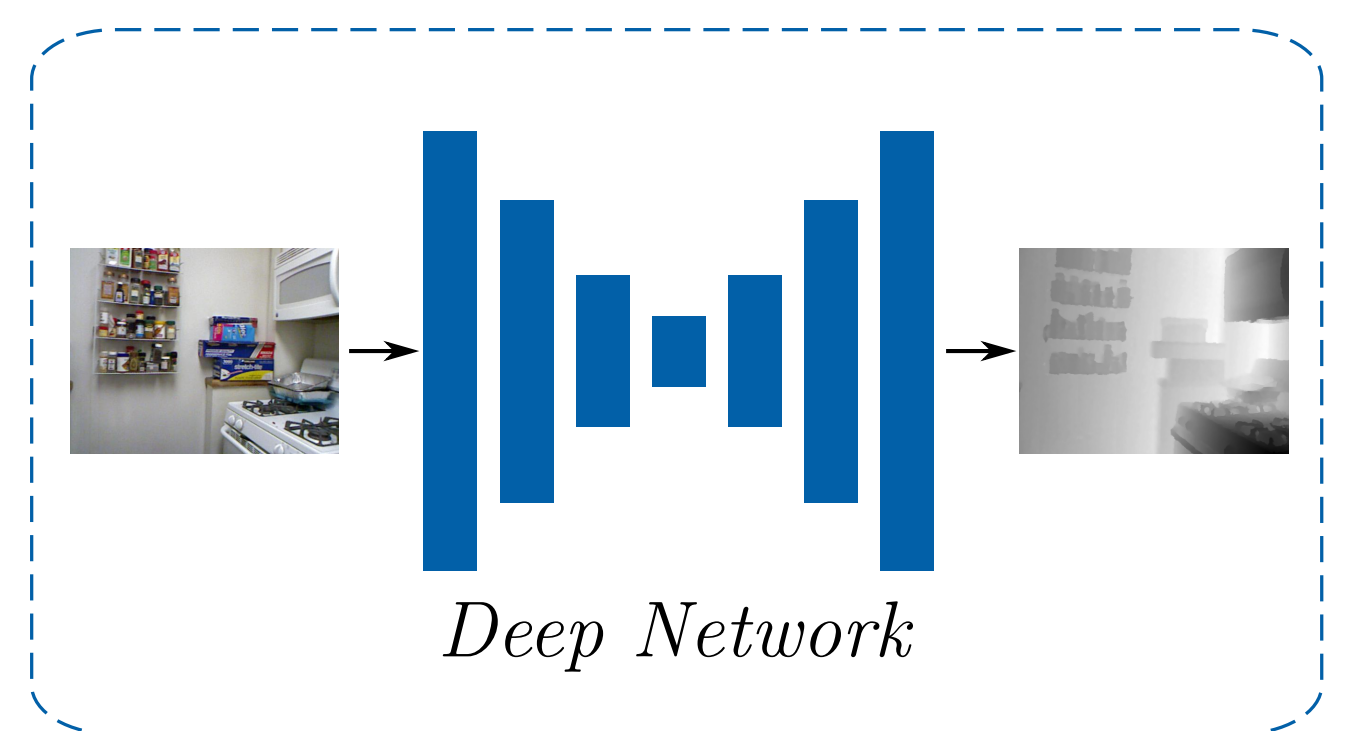
Inconvénients

- Temps de calcul important (~1fps).
- Manque de robustesse : méthode basée sur un modèle de la scène (*model-driven*).

Deep Learning

Permet une **abstraction de haut niveau** d'un jeu de données pour la modélisation.

Plusieurs **travaux antérieurs** pour l'estimation de profondeur : Eigen et al. [2] (**architecture multi-échelle**); Liu et al. [3] (**Conditional Random Fields, CRF**); et Dumas et al. (**DFD et Restricted Boltzmann Machines**) [4].



Avantages

- Traitement, analyse et prédictions rapides.
- Méthode basée sur les données (*data-driven*).

Inconvénients

- Besoin de beaucoup de données.
- Calcul coûteux (entraînement).
- Manque de fondement théorique.

Architecture

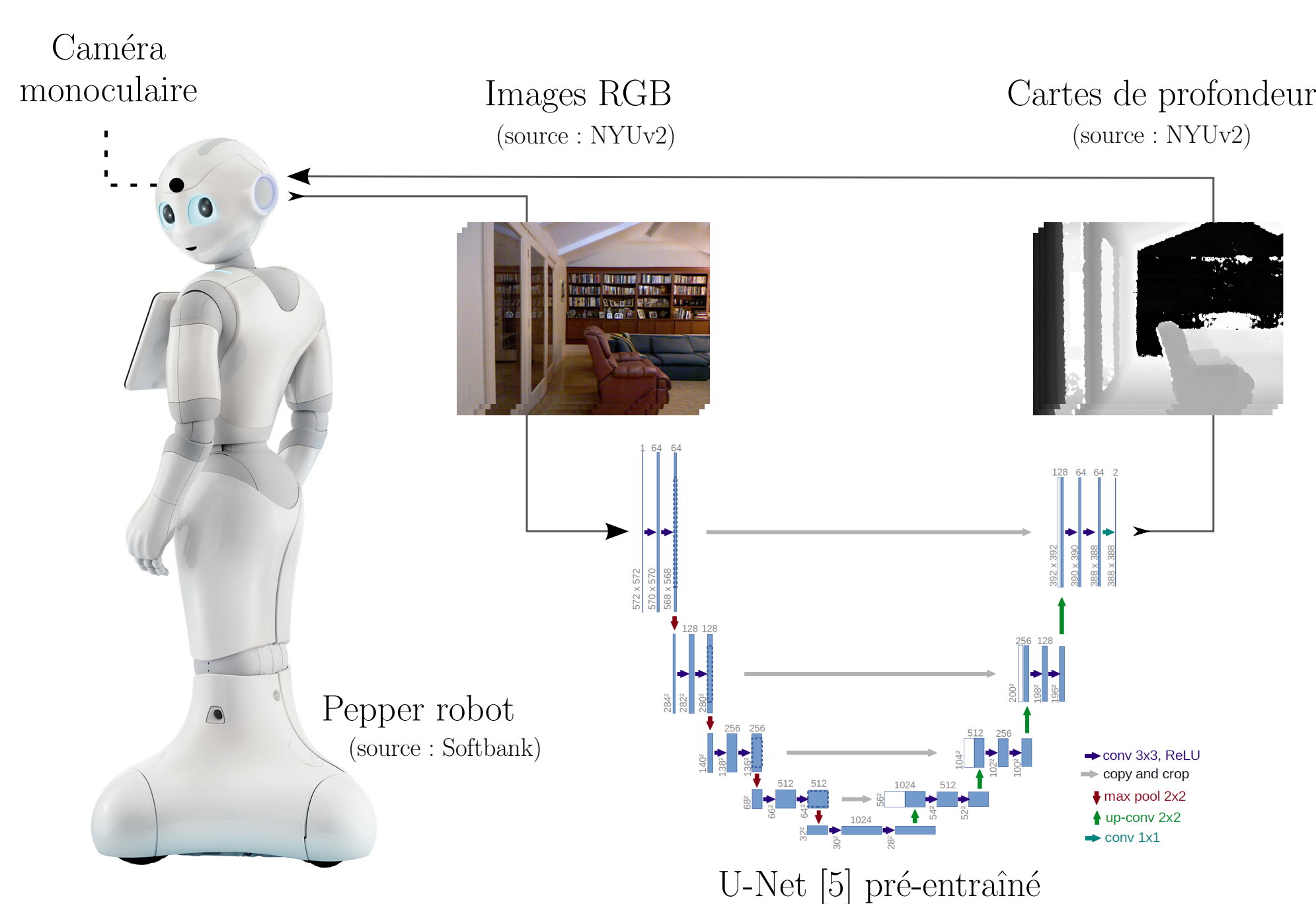
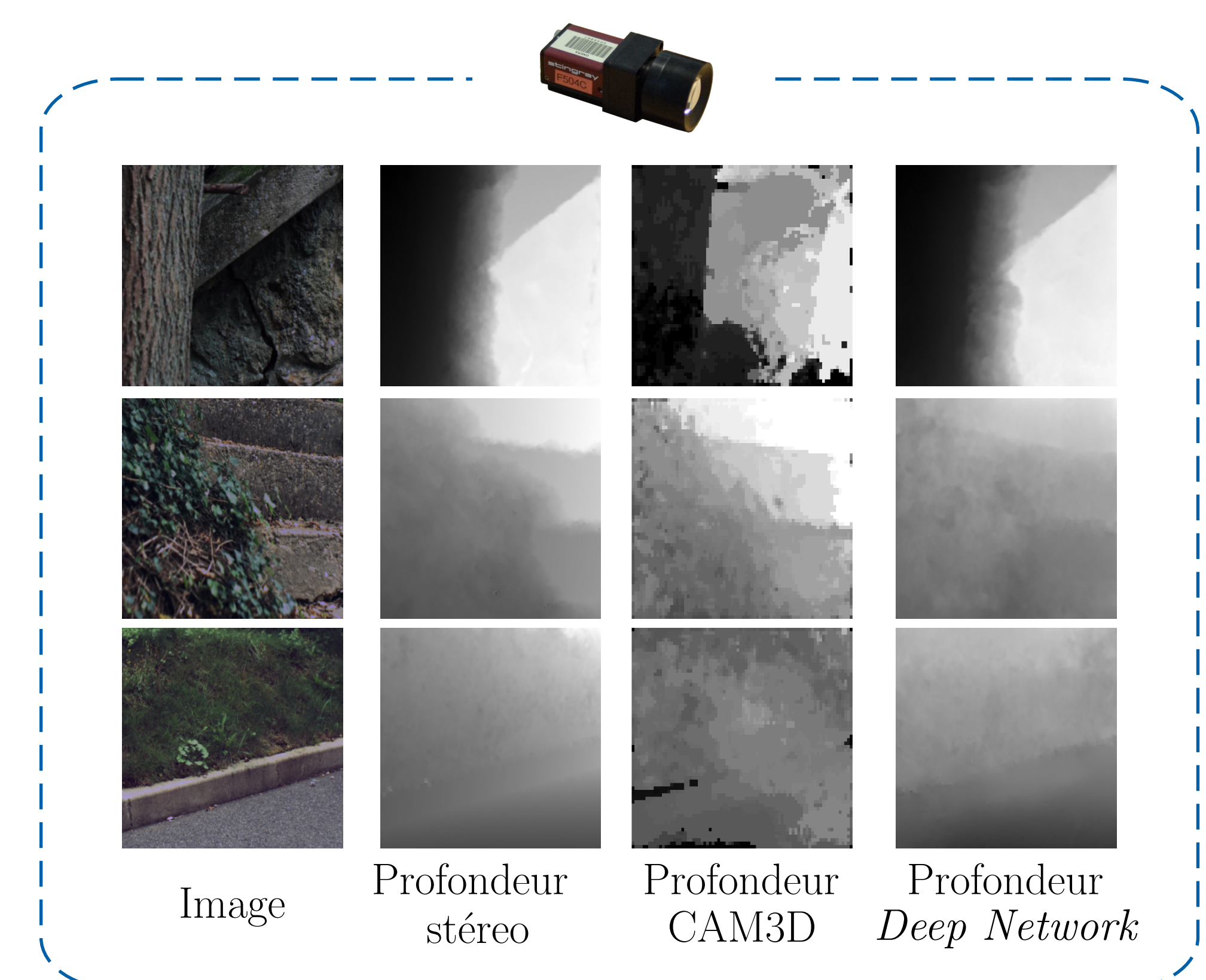
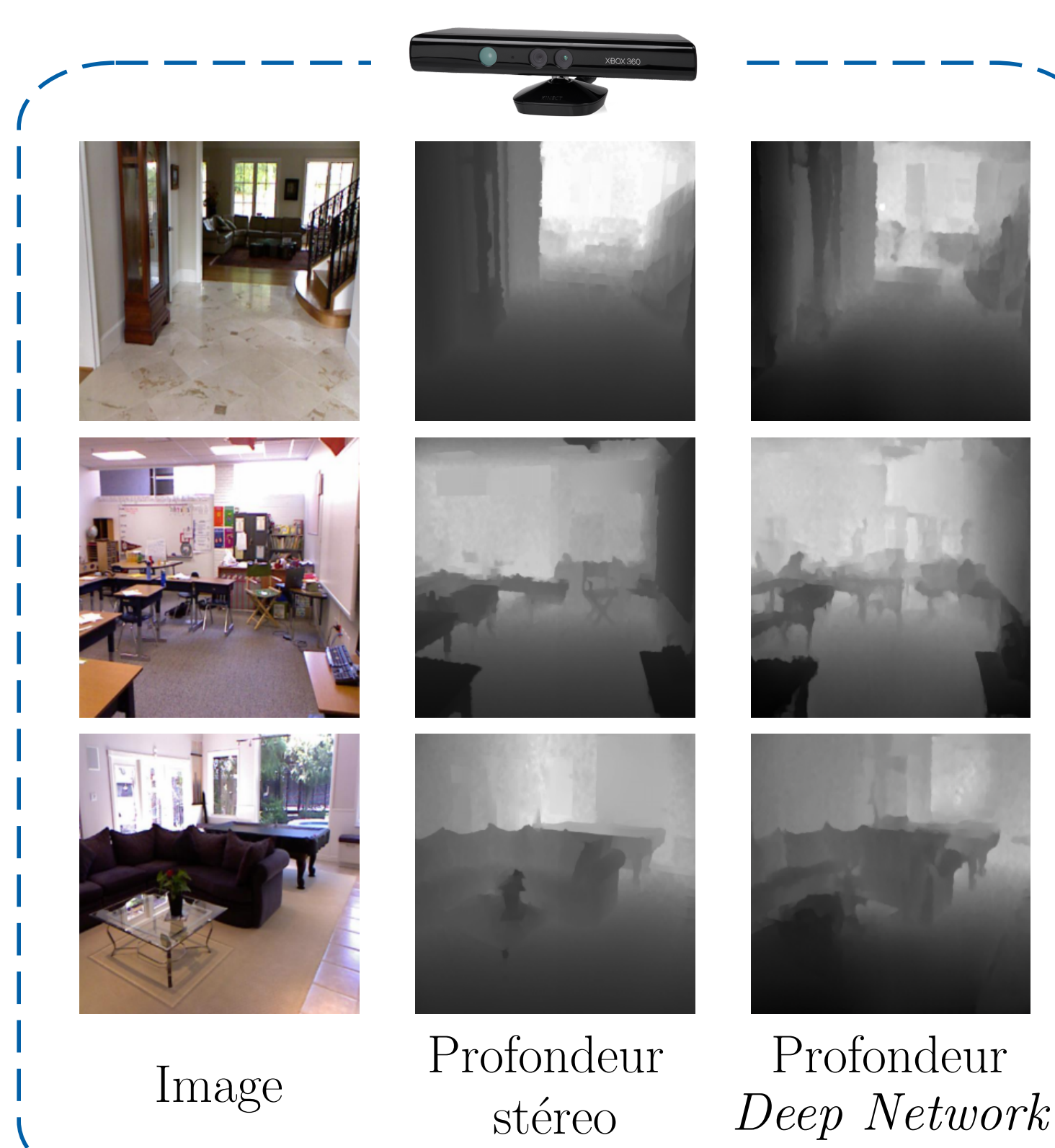


Fig. 1 - *Deep Learning* pour l'estimation de profondeur avec un robot porteur d'une caméra monoculaire.

Travaux en cours



Perspectives

1. Poursuite des travaux sur les **méthodes de deep learning** capables d'estimer automatiquement et rapidement la carte de profondeur à partir d'images d'un capteur dédié au DFD (nouvelle base de données, optimisation de l'architecture *Deep Learning*).
2. **Méthodes de restauration de la qualité des images**, afin de fournir une image couleur corrigée du flou de défocalisation (défloutée).
3. Réflexion sur la **conception conjointe optique/traitement** d'un imageur pour des applications 3D [1].

➔ **Intégration** des méthodes proposées dans le cadre des applications en **robotique**.

Bibliographie

- [1] P. Trouvé, "Conception conjointe optique/traitement pour un imageur compact à capacité 3D," Ph.D. dissertation, École Centrale de Nantes, 2012.
- [2] D. Eigen and R. Fergus, "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture," *ICCV*, 2016.
- [3] F. Liu, C. Shen, G. Lin, and I. D. Reid, "Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields," *IEEE TPAMI*, vol. 38, no. 10, p. 15, 2015.
- [4] T. Dumas, P. Trouvé-Peloux, and B. Le Saux, "Réseaux de neurones profonds pour estimer la profondeur grâce au flou de défocalisation," *Gretsi*, 2015.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net : Convolutional Networks for Biomedical Image Segmentation," *MICCAI*, 2015.