

PROPOSITION DE SUJET DE THÈSE

Intitulé : **Caméra 3D par Depth from Defocus et Machine Learning**

Référence : M TIS-DTIM-2016-007
(à rappeler dans toute correspondance)

Laboratoire d'accueil à l'ONERA :

Branche : TIS	Lieu (centre ONERA) : Palaiseau
Département : DTIM	
Unité : EVF	Tél. 0180386585 / 0180386573
Responsable ONERA : Pauline Trouvé-Peloux, Bertrand le Saux	Email : pauline.trouve@onera.fr ; bertrand.le_saux@onera.fr

Directeurs de thèse :

Nom : Andrès Almansa (Télécom Paristech) & Frédéric Champagnat (ONERA-DTIM)

Adresse :

Tél. : Email : andres.almansa@telecom-paristech.com ; frederic.champagnat@onera.fr

Sujet : **Caméra 3D par Depth from Defocus et Machine Learning**

(English version below)

Mots-clés : Vision 3D, machine learning, deep learning, depth from defocus

L'objectif de cette thèse est de développer de nouvelles approches d'apprentissage profond (ou deep learning) dans un contexte de caméra 3D monovoie passive utilisant le flou de défocalisation pour estimer la profondeur, approche désignée par le terme *Depth from Defocus (DFD)*. Comme l'estimation de la profondeur se fait au prix d'une dégradation de la qualité des images à cause du flou de défocalisation, l'enjeu du DFD est de disposer de méthodes rapides et précises pour l'estimation de flou et la restauration des images. Les approches par apprentissage promettent des temps de calcul réduits et une robustesse importante vis-à-vis de la variabilité de la scène par rapport à l'état de l'art. Ces nouvelles approches seront validées expérimentalement sur des optiques classiques et non conventionnelles, dédiées au DFD. La thèse permettra de construire un système de vision 3D compact et rapide, utilisable par exemple dans la boucle de perception d'un robot, pour la surveillance et l'exploration de sites.

Contexte

Depuis sa généralisation massive au cours des 10 dernières années, l'apprentissage profond (ou *deep learning*) a révolutionné la vision par ordinateur. Bien que les réseaux de neurones soient connus depuis longtemps, la création et l'entraînement de modèles comportant un nombre inégalé de couches a été rendu possible par le développement des moyens de calcul parallèles (GPU) et la création de bases de données d'envergure. La vision 3D peut également bénéficier de ces approches pour atteindre des objectifs de rapidité, de compacité et de coût qui permettront une rupture technologique. C'est sur cette idée que repose la thèse proposée dans ce dossier.

Les caméras 3D les plus répandues aujourd'hui reposent sur des approches actives telles que la technologie de mesure du temps de vol (ou *time of flight*) ou la projection de textures infra-rouge. Cependant ce type de caméra est réservé à une utilisation en intérieur en raison du capteur infra-rouge pouvant être perturbé par le rayonnement du soleil. De plus, ces caméras ont un encombrement et souvent un coût matériel qui empêchent leur utilisation dans des domaines tels que la robotique ou l'imagerie grand public. Le *Depth from Defocus* (DFD), reposant sur une estimation du flou de défocalisation d'une image pour en déduire la profondeur, permet d'avoir une information 3D sur la scène à partir d'un seul capteur passif. C'est donc une approche très prometteuse qui permet une utilisation en intérieur et extérieure avec une seule caméra. L'ONERA a développé sur le principe du DFD une caméra 3D monoculaire et passive : CAM3D, conçu spécifiquement pour le , mais dégradant en contrepartie encore plus la qualité de l'image par rapport à une caméra conventionnelle. Par ailleurs, les algorithmes de DFD non supervisés de l'état de l'art pâtissent encore aujourd'hui d'un temps de calcul non négligeable et d'un manque de robustesse vis-à-vis de la variabilité de la scène, en dehors du modèle de scène utilisé. Ces limitations motivent aujourd'hui l'exploration de nouvelles approches algorithmiques.

Cette thèse a pour objet de développer des méthodes rapides pour l'estimation de cartes de profondeur précises par DFD et la restauration d'images de grande profondeur de champ, le tout à l'aide d'un apprentissage profond. Ceci permettra de construire un système de vision 3D compact et rapide, utilisable dans la boucle de perception d'un robot, pour la surveillance et l'exploration de sites.

Objectifs scientifiques et résultats attendus

Les objectifs de la thèse sont les suivants :

- proposer des méthodes de deep learning capables d'estimer automatiquement et rapidement la carte de profondeur sans passer par une inversion analytique de la fonction de flou de l'optique. Pour cela, le doctorant devra mener à bien un travail théorique sur le choix du réseau de neurones le plus approprié, mais également expérimental de constitution de la base de données. Les résultats attendus sont les algorithmes d'estimation de profondeur, leur validation expérimentale sur différents types d'optique (conventionnelle ou non conventionnelle), en comparaison avec l'état de l'art.
- proposer des méthodes de restauration de la qualité des images, afin de fournir une image couleur défloutées. En particulier on cherchera à généraliser au cadre de la déconvolution d'un flou spatialement variable (DFV) comme celui d'une caméra 3D par DFD les techniques de régularisation non-locales qui ont révolutionné une large famille de problèmes inverses, qui n'inclut pas pour l'instant la DFV. On cherchera également à généraliser à la restauration d'images et plus particulièrement à la DFV la régularisation par réseaux de neurones, un nouveau champ de recherche qui a commencé très récemment à donner des résultats prometteurs dans le cadre du débruitage et la sur-résolution .
- contribuer à la réflexion sur la conception conjointe optique/traitement d'un imageur pour des applications 3D. Cela consistera notamment à mesurer l'influence des paramètres optiques sur la performance des méthodes d'apprentissage et d'amélioration de la qualité image. Le travail portera alors sur l'élaboration d'un modèle de performance permettant d'envisager l'optimisation de l'ensemble de la chaîne optique/traitement. Le résultat attendu est un nouveau modèle de performance permettant de prévoir de manière théorique la précision d'estimation de profondeur et la qualité image relative à un imageur, en fonction de ces paramètres optiques et traitements.

Compétences requises pour le candidat

Cette thèse nécessite de solides connaissances en traitement du signal et des images avec une spécialisation dans l'un ou plusieurs des domaines suivants : Apprentissage, Vision par ordinateur, Informatique et Optique.

La répartition du temps du doctorant envisagée serait de 1 à 2 jours par semaine au LTCI (Télécom ParisTech) et 3 à 4 jours par semaine à l'ONERA. Cette répartition permettra à tous les encadrants de suivre de près le travail de l'étudiant et lui donnera accès aux activités collectives des deux

laboratoires, ainsi que des échanges avec les chercheurs et doctorants de chaque laboratoire.

English version :

Keywords : *3D Monocular Camera; Depth From Defocus; Machine-learning; Neural Networks; Environment perception and recognition; Embedded intelligence*

The aim of this PhD is to develop new deep learning approaches for a single-lens 3D camera based on depth-from-defocus (DFD). Since DFD estimates depth thanks to defocus blur embedded in the optics, the challenge consists in having fast and precise methods for both blur estimation and image restoration. Deep-learning-based methods to estimate depth will allow constrained computation times and robustness to scene statistics variance. These new methods will be assessed experimentally on standard and unconventional optics. A practical objective will be to build a fast and compact 3D-vision system, for robot perception or scene monitoring and exploration.

1 Context, goals and applications

“Depth from Defocus” is a monocular passive depth estimation technique that is effective indoor as well as outdoor. Its compactness will allow to add 3D perception capacity to robots or everyday devices.

In a nutshell, Depth from defocus is based on estimation of the local blur due to the camera optics : two objects at different distances from the camera focal plane will have different blur levels, such that blur estimation allows to estimate distance. CAM3D outputs are a color image coupled with a depth map of the observed scene.

2 Objectives

- Propose deep learning methods able to estimate automatically the depth maps, using various flavours of convolutional neural networks (including fully-convolutional nets).
- Propose image restoration methods able to produce unblurred color images, using non-local regularization techniques used for solving inverse problems. Neural network-based regularization, a new technique for denoising and superresolution, will also be investigated
- Investigate new ways for joint conception of optics and processing for 3D applications

The PhD thesis objective will be to generate depth maps without analytical inversion of the blur function in tractable times, using machine-learning approaches:

- Efficient machine learning algorithms which are able to estimate depth from a single image using natural blur information. Several approaches will be investigated and combined, such as classification schemes based on deep neural networks and in particular convolutional networks [5] and recurrent neural networks able to learn sequence-to-sequence models.
- Representation spaces which are efficient to discriminate data while being invariant to translations and rotations, such as the scattering transform [7] or local Gaussian models [8]
- computational photography methods which increase the blur sensibility to depth for a better depth estimation.

Publication in the highest impact conferences and journals in computer vision and statistical machine learning is expected.

3 Applicant Profile

You should hold a an exceptionally strong academic record with an excellent degree (M.Sc., M.Eng. or equivalent) in Computer Science, Mathematics, or a related field (e.g. Electrical Engineering). Strong mathematical skills and a background in (at least) one of the following fields: machine learning, computer vision, optics and programming.

Collaborations extérieures : Télécom ParisTech

PROFIL DU CANDIDAT

Formation :

Master ou diplôme d'ingénieur avec excellents résultats

Spécificités souhaitées :

Solides connaissances en traitement du signal et des images avec une spécialisation dans l'un ou plusieurs des domaines suivants : Apprentissage, Vision par ordinateur, Informatique et Optique.