

Modèle photométrique RVB fin pour la co-conception optique/réseau de neurones

Marius DUFRAISSE¹ Pauline TROUVÉ-PELOUX¹ Jean-Baptiste VOLATIER² Frédéric CHAMPAGNAT¹

¹DTIS, ONERA - Université Paris-Saclay, F-91123, Palaiseau, France

²DOTA, ONERA - Université Paris-Saclay, F-91123, Palaiseau, France

Résumé – Les méthodes de co-conception optimisent une optique et un traitement mais résument généralement l’influence de l’optique à la réponse impulsionnelle du système sans considérer les changements d’intensité lumineuse qui modifient le rapport signal à bruit de l’image numérisée. Nous proposons un modèle de capteur imageur prenant en compte les différentes sources de bruit. Il est destiné à la co-conception de systèmes optiques et de réseaux de neurones. Nous appliquons cette modélisation à des exemples d’extension de profondeur de champ dans des cas réalistes : avec des images RVB et des objets hors de l’axe optique.

Abstract – Co-design methods generally model the influence of the optical parameters only on the point spread function of the system without considering changes in light intensity that modify the signal to noise ratio of the digitized image. We include multiple noise sources in a model of imaging sensors intended for the co-design of an optical system and a neural network. We use it in co-design examples for extended depth of field in a realistic setting : with RGB images and off-axis objects.

1 Introduction

Les systèmes d’imagerie sont composés d’un objectif, assemblage de plusieurs lentilles, et d’un traitement. Ces éléments sont généralement conçus séparément : l’objectif est optimisé pour minimiser les aberrations et le traitement mis au point pour une tâche spécifique en faisant abstraction du matériel ayant produit les images traitées. Les méthodes de co-conception ont été introduites pour réduire cet écart et optimiser simultanément ces deux composants [1]. Récemment, il a été montré que l’on peut optimiser conjointement les paramètres d’un réseau de neurones et d’une lame de phase placée sur le diaphragme d’un objectif existant [2, 3]. Ainsi, l’extension de profondeur de champ, la super-résolution ou l’estimation de profondeur peuvent être améliorées. La co-conception peut aussi être utilisée pour optimiser plusieurs lentilles et un réseau de neurones [4, 5, 6, 7]. L’optimisation conjointe d’un réseau de neurones et d’une optique repose sur une simulation dérivable du système optique à l’aide de l’optique de Fourier ou du tracé de rayons. Ces modèles simulent l’influence des paramètres optiques sur les aberrations en calculant la réponse impulsionnelle du système mais ne prennent pas en compte leurs effets sur l’intensité lumineuse.

Pourtant, certains paramètres optiques, et notamment l’ouverture, ont une influence directe sur le flux lumineux traversant l’objectif : fermer le diaphragme diminue les aberrations mais abaisse la quantité de lumière qui atteint le capteur au détriment du bruit. Le meilleur compromis possible dépendra de la capacité du traitement à débruiter ou à déconvoluer et il est donc intéressant d’intégrer l’optimisation de l’ouverture à la co-conception pour trouver sa valeur optimale.

Cela a d’abord été fait en considérant l’ouverture comme un hyper-paramètre et en testant plusieurs valeurs [8]. Pour l’optimiser directement, nous avons déjà montré comment lier le bruit de Schottky aux paramètres optiques [9] plutôt que de simplement ajouter un bruit gaussien de variance fixe comme le font la plupart des travaux en co-conception.

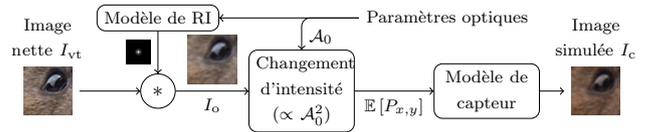


FIGURE 1 : Chaîne de génération d’image : l’image de vérité terrain I_{vt} est convoluée avec la RI pour obtenir l’image optique I_o , puis un changement de la dynamique en fonction du rayon de la pupille d’entrée \mathcal{A}_0 est appliqué avant de simuler l’image bruitée et numérisée par le capteur I_c .

Nous proposons ici une modélisation plus complète du capteur, proche de celle proposée dans [6] mais intégrée à la chaîne dérivable de simulation des images. Nous l’utilisons pour l’extension de profondeur de champ dans un cadre réaliste puisque nous considérons des images en couleur non dématricées et des objets hors de l’axe optique.

2 Méthode

Nous proposons une chaîne de formation d’images simulant de façon dérivable la réponse impulsionnelle d’une optique, sa radiométrie, le bruit et les non linéarités introduites par le capteur photographique. La Figure 1 illustre la chaîne de simulation proposée.

2.1 Simulation dérivable de l’optique

L’effet d’un objectif sur une image est décrit par sa réponse impulsionnelle (RI) : une image nette I_{vt} forme une image $I_o = I_{vt} * RI$ lorsqu’elle est projetée sur le capteur par l’objectif. Plusieurs modèles permettent de calculer la RI et sa jacobienne par rapport aux paramètres optiques [10].

Afin d’optimiser des optiques composées de lentilles épaisses et avec des objets éloignés de l’axe optique nous

utilisons une bibliothèque de tracé de rayons dérivable (TRD) [4]. Elle utilise les lois de Snell-Descarte pour calculer la trajectoire de rayons lumineux au travers d'une succession de lentilles et calculer les gradients de ces trajectoires par rapport aux paramètres optiques.

Le TRD permet donc de tracer des rayons émis par un point objet donné pour qu'ils échantillonnent uniformément la pupille et d'obtenir les intersections de ces rayons avec le plan du capteur et donc une approximation discrète de l'énergie reçue par chaque pixel. Pour obtenir la valeur de la RI en chaque pixel et tenir compte de la diffraction, nous intégrons la convolution entre la distribution discrète obtenue par TRD et une tâche d'Airy [5]. En effet, réduire l'ouverture réduit les aberrations mais peut introduire de la diffraction, en convoluant la distribution avec une tâche d'Airy dont le rayon varie avec celui de la pupille d'entrée il n'est pas possible d'atteindre des cas irréalistes où la RI serait ponctuelle. Le diamètre de la pupille d'entrée \mathcal{A}_0 est calculé de façon dérivable à l'aide des outils de TRD. L'intégration est effectuée en utilisant la méthode des rectangles avec une grille d'échantillonnage de 4×4 points par pixel.

2.2 Flux lumineux

La RI étant normalisée il faut multiplier I_o par le flux F traversant l'objectif pour modéliser les changements d'intensité introduits par l'ouverture. Pour une source de luminance L_S et d'aire A_S , avec une pupille d'entrée de rayon angulaire α , ce flux est $F = \pi L_S A_S \sin^2(\alpha)$ [11]. Pour un objet à distance d , chaque pixel est éclairé par une surface d'aire $A_S = \frac{t_{pix}}{M_d}$, avec t_{pix} la taille des pixels et M_d le grossissement linéaire à une distance d ; ici $M_d \approx \frac{f}{d}$ avec f la longueur focale. Tout comme f , la pupille d'entrée \mathcal{A}_0 est calculée à l'aide du TRD et vérifie $\sin \alpha \approx \frac{\mathcal{A}_0}{d}$ car $d \gg \mathcal{A}_0$. Ainsi, l'énergie atteignant un pixel (x, y) en un temps Δ_t est

$$E(x, y) = \Delta_t \pi L_S I_0(x, y) \left(\frac{\mathcal{A}_0 t_{pix}}{f} \right)^2. \quad (1)$$

2.3 Modèle de capteur

Nous avons précédemment proposé un modèle simplifié de capteur ne comportant que le bruit de Schottky qui dépend directement de l'intensité de l'image et donc de l'ouverture [9]. Nous complétons ici ce modèle avec d'autres sources de bruit et de non linéarités [6].

L'énergie $E(x, y)$ atteignant un pixel est convertie en un nombre moyen de photons $\mathbb{E}[P_{x,y}]$ atteignant le capteur à l'aide de l'équation de Planck. Le nombre de photons recueilli par le capteur suit une loi de Poisson de moyenne $\mathbb{E}[P_{x,y}]$. Afin d'obtenir une implémentation dérivable, le nombre de photons captés par chaque récepteur est tiré aléatoirement selon une loi normale de moyenne et variance $\mathbb{E}[P_{x,y}]$ qui approxime cette loi de Poisson.

Pour chaque longueur d'onde considérée, l'efficacité quantitative permet de calculer le nombre d'électrons obtenus par le capteur photosensible. Le bruit thermique est simulé en ajoutant à chaque pixel un bruit suivant une loi de Poisson de paramètre $\mu_{ref} \Delta_t$, la température étant supposée constante. Le bruit de lecture est modélisé par une loi normale de paramètres μ_l, σ_l^2 . Nous n'optimisons pas les paramètres de ces bruits, leur implémentation n'a donc pas à être dérivable.

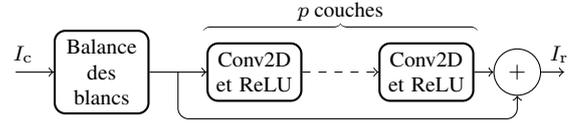


FIGURE 2 : Le réseau utilisé, un seul hyper-paramètre varie dans nos expériences : le nombre de couches p .

Chaque capteur peut recueillir au plus C électrons avant d'être saturé, nous appliquons donc un seuil quand cette limite est atteinte. Un gain est ensuite appliqué pour passer du nombre d'électrons à une valeur numérique normalisée. La numérisation du signal introduit une quantification, nous ajoutons donc un bruit de quantification uniforme sur $[0, 2^{-b}]$ pour simuler la précision limitée du signal numérisé, b est le nombre de bits de résolution du convertisseur. Pour simuler l'acquisition d'images sur plusieurs canaux, un filtre de Bayer est appliqué pour obtenir des images brutes de capteur.

2.4 Traitement

Notre modèle de capteur applique un filtre de Bayer pour obtenir une image RGGB. Tous les traitements étudiés commencent donc par un filtre d'interpolation linéaire fixe pour le dématricage. Ensuite nous étudions un filtrage soit par une seule convolution dont on fera varier la taille du noyau ; soit par un réseau résiduel simple ayant p couches précédées par une balance des blancs optimisée pendant l'entraînement (voir Figure 2).

3 Expériences

Nous proposons plusieurs expériences dans lesquelles nous appliquons notre modèle de formation des images à une tâche d'amélioration de la qualité image. En effet, les réseaux de neurones ayant la possibilité de déconvoluer ou de débruiter, leur optimisation conjointe avec l'ouverture peut conduire à deux modes de fonctionnement distincts. Dans un premier temps nous étudions un système avec des objets sur l'axe à une distance fixe, puis sur une plage de distances étendue, et enfin avec en plus des objets hors de l'axe optique.

Les simulations sont réalisées avec un doublet de Gauss de focale 100mm optimisé avec Zemax pour avoir une distance de mise au point de 25m. Le rayon du diaphragme \mathcal{A} et la position du capteur seront alors optimisés à partir de 9.95mm et 133.3mm. Pour les paramètres du capteurs nous utilisons des valeurs réalistes [6] : $\Delta_t = 0.02s$, $\mu_{ref} = 25e^{-s^{-1}}$, $\mu_l = 3.25$, $\sigma_l^2 = 16$, $C = 35000e^{-}$, et $b = 12$. L'optique et le traitement sont optimisés pour minimiser la norme L_1 entre la vérité terrain et la reconstruction sur des patches de taille 64×64 issus du Describable Texture Dataset [12]. Nous utilisons Adam avec des pas de 10^{-4} pour le traitement, 10^{-3} pour la position du capteur, et 10^{-2} pour le rayon de l'ouverture.

3.1 Objet à distance fixe

Nous proposons tout d'abord un exemple simple de restauration d'image avec un objet situé à 27m et une mise au point à 25m. Dans cet exemple le rayon du diaphragme et le traitement sont optimisés. La figure 3 donne l'écart-type du spot

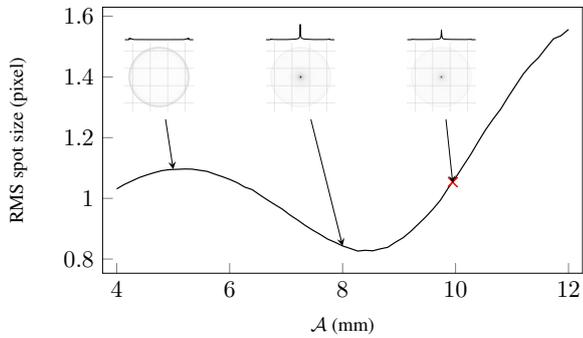


FIGURE 3 : Valeur de RMS spot size en fonction de l’ouverture pour un système mis au point à 25m et des objets à 27m. Images et profils de la RI pour plusieurs ouvertures.

TABLE 1 : Optimisation de l’ouverture et d’un réseau de restauration pour des objets à une distance fixe.

Traitement	Intensité	\mathcal{A}^* (mm)	PSNR (dB)
Balance des blancs	$\times 1$	10.5	23.6
Conv. 3×3	$\times 1$	9.3	24.2
Conv. 3×3	$\times 0.1$	11.3	21.8
Conv. 3×3	$\times 0.01$	12.0	16.5
Conv. 7×7	$\times 1$	10.3	24.5
Réseau 3c.	$\times 1$	8.9	27.40
Réseau 7c.	$\times 1$	9.3	27.7
Réseau 7c.	$\times 0.1$	9.9	24.5
Réseau 7c.	$\times 0.01$	10.6	20.7

diagramme en fonction de l’ouverture ainsi que des exemples de RI. Pour de petites ouvertures, la forme de la RI est typique d’un flou de défocalisation. En ouvrant, l’aberration sphérique concentre les rayons près du centre et l’étalement de la RI est moins marqué. Dans ce cas, l’énergie est concentrée dans ce pic central et l’écart-type du spot est plus faible : il n’est donc pas croissant pour les objets hors de la profondeur de champ.

Nous optimisons d’abord l’ouverture avec uniquement une balance des blancs. Les résultats obtenus sont affichés dans la Table 1. L’ouverture optimisée \mathcal{A}^* est proche de l’ouverture initiale avec laquelle l’objectif avait été optimisé sans prendre en compte le traitement. Dans ce cas le PSNR est de 23.6dB.

Les traitements plus complexes améliorent la qualité image : les convolutions permettent de gagner 1dB de PSNR et les réseaux 3dB. Lors de ces expériences l’ouverture reste proche de 10mm contrairement à nos expériences précédentes [9] dans lesquelles elle diminuait vers 8mm. Cela vient des paramètres choisis ici pour s’approcher de ceux d’un capteur standard et qui imposent une luminosité de la scène plus faible : fermer l’ouverture pour atteindre le minimum de RMS spot size (voir Figure 3) introduirait ici trop de bruit.

La Table 1 donne aussi les résultats obtenus avec une intensité lumineuse plus faible. Lorsqu’elle diminue, l’optimisation augmente l’ouverture pour limiter le bruit ce qui est un signe de la validité du modèle proposé. Cette augmentation de l’ouverture est moins marquée lorsque le traitement est un réseau de neurones ce qui suggère que ceux-ci sont plus adaptés pour débruiter qu’une unique convolution.

TABLE 2 : Optimisation de l’ouverture, de la position du capteur, et d’un réseau de restauration pour des objets entre 20m et 40m, sur et hors de l’axe optique avec un modèle de capteur de plus en plus complet.

Changement	\mathcal{A}^* (mm)	Point (m)	PSNR (dB)
Référence [9]	3.7	30.0	28.9
Intensité $\times 1/15$	9.9	24.0	28.0 (-0.9)
Bruits électroniques et efficacité quantique	10.1	23.5	27.8 (-0.2)
Saturation	10.1	23.5	27.7 (-0.1)
Bruit de quantification	9.8	26.0	27.3 (-0.4)
Image RVB	10.4	24.0	28.8 (+1.5)
Filtre de Bayer	10.2	24.0	27.1 (-1.7)
Hors-axe	10.0	23.0	25.2 (-1.9)

3.2 Profondeur de champ étendue

Nous considérons maintenant des objets placés entre 20m et 40m et optimisons l’ouverture, la position du capteur, et le traitement pour améliorer la qualité image. La Table 2 présente les rayons de diaphragme, la distance de mise au point et le signal rapport à bruit obtenus en optimisant l’ouverture, la position du capteur et le traitement (réseau à 7 couches) pour différentes modélisations du capteur de plus en plus complexes à partir d’un modèle avec seulement du bruit de Schottky et une seule longueur d’onde [9] qui atteint 28.9dB.

Comme dans la section 3.1, nous reprenons le paramétrage des bruits de [6] ce qui correspond à une diminution de l’intensité par un facteur 15 par rapport à [9]. Cette baisse de la luminosité entraîne une augmentation de l’ouverture pour conserver un rapport signal à bruit suffisant sans que la qualité image ne se dégrade trop.

Les bruits thermiques et de lecture ainsi que l’efficacité quantique diminuent assez peu la performance du système (-0.2 dB de PSNR) car la scène est suffisamment lumineuse pour que ces sources de bruits soient faibles devant le nombre de photons récoltés. L’ajout d’un bruit de quantification a un impact assez faible sur la qualité du système optimisé. Toutefois, nous avons fixé le nombre de bits à 12 ce qui est plutôt élevé, avec seulement 8 bits l’optimisation donne une système ayant un PSNR de 26.1dB.

Le passage à un système fonctionnant avec trois canaux améliore les performances du système de 1.5dB. À cause du chromatisme, les trois longueurs d’ondes correspondent à 3 distances de mise au point entre 23.5m et 28m qui permettent d’améliorer la profondeur de champ en réalisant du transfert de fréquence [13]. Ici la flexibilité des réseaux de neurones permet d’exploiter le chromatisme de l’optique pour améliorer les performances globales du système : l’utilisation de réseaux de neurones en co-conception favorise l’émergence d’optiques non conventionnelles car ils n’imposent pas d’a priori sur le mode de fonctionnement du traitement. L’ajout d’un filtre de Bayer réduit significativement la qualité des résultats obtenus. Cela pourrait en partie être corrigé en optimisant les paramètres du dématricage avec ceux de l’optique et du réseau de neurones. Par ailleurs, avec une ouverture initiale de 7.5mm ou 12mm

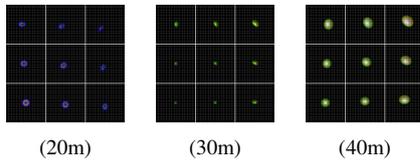


FIGURE 4 : RI du système obtenu avec un traitement constitué d'un réseau à 3 couches pour des angles de 0° , 5° et 10° . Pour chaque RI le quadrillage fin indique les pixels du capteur.

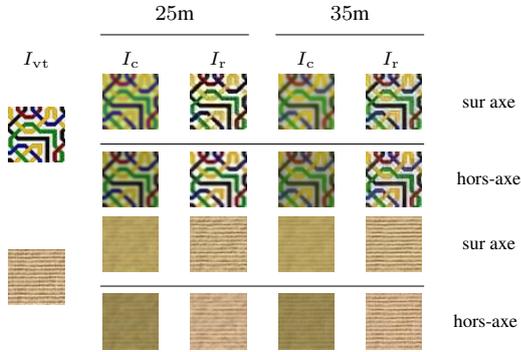


FIGURE 5 : Exemples de patches sur l'axe obtenus à deux distances avec les systèmes optimisés sur et hors de l'axe optique. I_{vt} est l'image de référence, I_c celle numérisée par le capteur et I_r le résultat du traitement.

celle-ci converge encore vers une valeur comprise entre 9mm et 10mm.

Enfin, nous optimisons une optique et un unique réseau pour des objets dans un champ angulaire de 10° . Les paramètres optiques obtenus dans ce cas sont similaires à ceux de l'optimisation sur l'axe. Toutefois, les aberrations présentes dans les RI hors-axe (voir Figure 4) sont plus difficiles à corriger et les performances du système en pâtissent significativement : le PSNR diminue de 2dB. Cette baisse est perceptible comme on peut le voir sur la Figure 5 qui présente des exemples de patches obtenus avec le système optimisé sur l'axe et celui hors-axe. Dans ce dernier cas, les résultats sont plus flous à cause de la variation de la RI dans le champ. Un traitement variable dans le champ pourrait permettre d'améliorer ces résultats. L'entraînement dure environ 5h, il est dominé par le TRD.

4 Conclusion

Nous proposons un modèle de capteur complété et lié aux paramètres optiques. Il s'intègre dans la chaîne de simulation dérivable utilisée pour la co-conception optique-réseau de neurones. Il prend en compte les effets des paramètres optiques sur la dynamique de l'image et donc sur le bruit et la saturation de celle-ci. Nous appliquons ce modèle dans un cadre réaliste d'extension de profondeur de champ, avec des images monochromes, mais aussi en couleur et avec des objets hors de l'axe optique. Ces expériences illustrent l'intérêt des réseaux de neurones pour la co-conception : leur polyvalence autorise l'émergence de modes de fonctionnement originaux.

Enfin, nous allons continuer à utiliser ce modèle dans nos travaux de co-conception notamment pour l'appliquer à des tâches plus complexes de vision par ordinateur comme la segmentation sémantique.

Références

- [1] W. T. CATHEY et E. R. DOWSKI : New paradigm for imaging systems. *Appl. Opt.*, 41(29):6080–6092, Oct 2002.
- [2] V. SITZMANN, S. DIAMOND, Y. PENG, X. DUN, S. BOYD, W. HEIDRICH, F. HEIDE et G. WETZSTEIN : End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph.*, 37(4), jul 2018.
- [3] H. HAIM, S. ELMALEM, R. GIRYES, A. M. BRONSTEIN et E. MAROM : Depth estimation from a single image using deep learned phase coded mask. *IEEE Transactions on Computational Imaging*, 4:298–310, 2018.
- [4] A. HALÉ, P. TROUVÉ-PELOUX et J.-B. VOLATIER : End-to-end sensor and neural network design using differential ray tracing. *Opt. Express*, 29(21):34748–34761, Oct 2021.
- [5] Z. LI, Q. HOU, Z. WANG, F. TAN, J. LIU et W. ZHANG : End-to-end learned single lens design using fast differentiable ray tracing. *Opt. Lett.*, 46(21):5453–5456, Nov 2021.
- [6] E. TSENG, A. MOSLEH, F. MANNAN, K. St. ARNAUD, A. SHARMA, Y. PENG, A. BRAUN, D. NOWROUZEHARAI, J.-F. LALONDE et F. HEIDE : Differentiable compound optics and processing pipeline optimization for end-to-end camera design. *ACM Transactions on Graphics (TOG)*, 40:1 – 19, 2021.
- [7] Q. SUN, C. WANG, Q. FU, X. DUN et W. HEIDRICH : End-to-end complex lens design with differentiable ray tracing. *ACM Trans. Graph.*, 40(4), jul 2021.
- [8] S. r. M. ROSTAMI, S. PINILLA, I. SHEVKUNOV, V. KATKOVNIK et K. EGIAZARIAN : On design of hybrid diffractive optics for achromatic extended depth-of-field (EDoF) RGB imaging, 2022.
- [9] M. DUFRAISSE, P. TROUVÉ-PELOUX, J.-B. VOLATIER et F. CHAMPAGNAT : Deblur or denoise : the role of an aperture in lens and neural network co-design. *Opt. Lett.*, 48(2):231–234, Jan 2023.
- [10] M. DUFRAISSE, P. TROUVÉ-PELOUX, J.-B. VOLATIER et F. CHAMPAGNAT : On the use of differentiable optical models for lens and neural network co-design. In *Unconventional Optical Imaging III*, volume 12136, pages 176 – 185. SPIE, 2022.
- [11] J.-L. MEYZONNETTE et T. LÉPINE : *Bases de Radiométrie Optique*. Cépaduès éditions, 2 édition, 2001.
- [12] M. CIMPOI, S. MAJI, I. KOKKINOS, S. MOHAMED, et A. VEDALDI : Describing textures in the wild. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [13] F. GUICHARD, H.-P. NGUYEN, R. TESSIÈRES, M. PYANET, I. TARCHOUNA et F. CAO : Extended depth-of-field using sharpness transport across color channels. In *Digital Photography V*, volume 7250, pages 192–203. SPIE, 2009.