



**HAL**  
open science

# Pixel Level Depth Estimation in Computer Generated Holograms

Nabil Madali, Antonin Gilles, Patrick Gioia, Luce Morin

► **To cite this version:**

Nabil Madali, Antonin Gilles, Patrick Gioia, Luce Morin. Pixel Level Depth Estimation in Computer Generated Holograms. ORASIS 2023, Laboratoire LIS, UMR 7020, May 2023, Carqueiranne, France. hal-04219354

**HAL Id: hal-04219354**

**<https://hal.science/hal-04219354>**

Submitted on 27 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Pixel Level Depth Estimation in Computer Generated Holograms

Nabil Madali<sup>1,2</sup>

Antonin Gilles<sup>1</sup>

Patrick Gioia<sup>1,3</sup>

Luce Morin<sup>1,2</sup>

<sup>1</sup> Institute of Research & Technology b<>com, Cesson-Sévigné, France

<sup>2</sup> Univ Rennes, INSA Rennes, CNRS, IETR – UMR 6164, F-35000 Rennes

<sup>3</sup> Orange Labs, Cesson-Sévigné, France

Auteur correspondant : nabil.madali@b-com.com

## Résumé

Dans cet article, nous présentons et examinons une méthode basée sur l'apprentissage profond pour l'extraction de cartes de profondeur à partir d'hologrammes générés par ordinateur. Cette méthode, qui se distingue des méthodes conventionnelles d'extraction de profondeur basées sur la mise au point également appelées Depth from Focus (DFF), propose une approche plus avancée et performante.

## Mots Clef

Imagerie 3D, holographie, estimation de la profondeur, apprentissage auto-supervisé, profondeur à partir de la mise au point.

## 1 Introduction

Les hologrammes générés par ordinateur [5] permettent de simuler des objets ou des scènes en 3D dans un environnement virtuel, créant une expérience immersive pour l'utilisateur. Avec leur capacité à imiter la réalité de manière convaincante, ils peuvent être utilisés dans divers domaines tels que l'art, l'architecture, la médecine, l'éducation, etc.

L'hologramme enregistré est une source riche d'informations sur la scène 3D acquise. Cependant, le manque de localisation spatiale [1] des objets de la scène à l'intérieur du plan d'hologramme peut être un obstacle à leur extraction. En effet, chaque onde lumineuse provenant de chaque point de la scène contribue à chaque pixel lors de l'enregistrement de l'hologramme.

Une manière relativement simple de retrouver la localisation spatiale et ainsi de réaliser un apprentissage plus localisé avec une utilisation de GPU plus faible est d'utiliser des reconstructions numériques [16]. Etant donné un hologramme  $H$  de taille  $L \times L$  et une distance de reconstruction  $z$ , une reconstruction numérique peut être calculée à l'aide du spectre angulaire [3] comme suit :

$$\mathcal{P}_z\{H\}(x, y) = \mathcal{F}^{-1} \left\{ \mathcal{F}(H) e^{j2\pi z \sqrt{\lambda^{-2} - f_x^2 - f_y^2}} \right\} (x, y), \quad (1)$$

où  $\mathcal{F}$  désigne la transformée de Fourier 2D,  $f_x$  et  $f_y$  sont les fréquences spatiales dans les directions horizontale et verticale du plan de l'hologramme, respectivement, et  $\lambda$  est la longueur d'onde d'acquisition.

Chaque reconstruction numérique est affectée par du bruit de speckle [13], à l'exception de certaines parties qui sont claires si elles se trouvent à la même profondeur que la distance de reconstruction. Ainsi, reconstruire la géométrie de la scène correspond à détecter, pour chaque pixel du plan de reconstruction  $(m, n)$ , la profondeur de reconstruction  $z_i$  à laquelle le niveau de netteté atteint localement son maximum en utilisant un patch centré sur le pixel  $R_{m,n}$  à l'aide d'un opérateur de mesure de netteté [9].

A cet effet, nous proposons d'utiliser un réseau neuronal dénoté  $\mathcal{G}$  basé sur l'architecture U-Net [10] pour segmenter les parties mises au point sur chaque reconstruction numérique effectuée. Le problème peut être formulé mathématiquement de la manière suivante :

$$\mathcal{L} = \sum_{u,v} \sum_{i=1}^N \mathbb{1}_{\{d(u,v)=z_i\}} \cdot \log \hat{I}_i(u, v) \quad (2)$$

$$\hat{I}_i = \mathcal{G}\{\Lambda_{p,q,i}\}, \quad (3)$$

$$\Lambda_{p,q,i} = |\mathcal{P}_{z_i}\{H\}|(ps + u, qs + v), \quad u \in [0, s - 1], \quad v \in [0, s - 1], \quad i \in [1, N] \quad (4)$$

où  $\mathbb{1}$  est la fonction indicatrice,  $\lfloor a \rfloor$  est la plus grande valeur entière inférieure ou égale à  $a$ ,  $H$  et  $W$  sont respectivement la hauteur et la largeur du volume de reconstruction,  $\hat{I}_i$  est la carte de mise au point prédite correspondant à la reconstruction holographique effectuée à la distance  $z_i$ ,  $s$  est la taille de patch utilisée, et  $z_i$  est la profondeur de reconstruction, donnée par

$$z_i = \frac{z_{\max} - z_{\min}}{N} i + z_{\min}. \quad (5)$$

où  $z_{\min}$  et  $z_{\max}$  sont les valeurs de profondeur les plus petites et les plus grandes possibles des points de la scène. Une fois l'apprentissage du réseau terminé, la profondeur prédite par pixel peut être extraite selon la formule suivante

$$\hat{d}_{p,q}(u, v) = \arg \max_{i \in [1, N]} \hat{I}_i(u, v) \quad (6)$$

Cette formulation permet de garantir que le réseau identifie un seul plan de focalisation optimal pour chaque volume de reconstruction d’entrée.

## 2 Résultats expérimentaux

Afin de former le réseau neuronal, un grand ensemble de données a été acquis, composé de 4200 hologrammes RGB obtenus à partir de trois scènes différentes (*Piano*, *Table*, *Woods*), en utilisant une méthode basée sur une décomposition de la scène en tranches [2]. La résolution était de  $1024 \times 1024$ , avec un pas de pixel de  $6\mu\text{m}$  et 1400 angles d’acquisition. Pour l’évaluation, deux ensembles de données ont été utilisés : un ensemble de test composé de 100 hologrammes additionnels pour chaque scène d’entraînement, et un ensemble de validation comprenant 200 hologrammes enregistrés à partir de deux scènes non vues (*Cars*, *Dices*) pendant la phase d’entraînement.

Les reconstructions holographiques ont été calculées à des profondeurs uniques à partir de la vérité terrain de la carte de profondeur pour chaque hologramme. Le volume de reconstruction obtenu a été décomposé spatialement en patches non chevauchants. Les patches extraits à la même position spatiale sont introduits dans le réseau, qui est entraîné à prédire les cartes de mise au point correctes en utilisant la perte d’entropie croisée comme supervision.

Le réseau a été formé pendant 200 époques en utilisant l’optimiseur Adam et une décroissance d’apprentissage avec un facteur gamma de 0,8 toutes les 10 époques et un taux d’apprentissage initial de 0,1. Le réseau a été comparé à neuf opérateurs de mise au point répertoriés dans le Tableau 1. Les opérateurs ont été appliqués sur des fenêtres de taille  $17 \times 17$  qui ont donné les meilleurs résultats lors des expériences précédentes.

Opérateur de mise au point	Abbr.	Opérateur de mise au point	Abbr.
Énergie du Laplacien [12]	LAPE	Laplacien modifié [7]	LAPM
Variance du Laplacien [8]	LAPV	Laplacien diagonal [14]	LAPD
Variance des coefficients d’ondelettes [15]	WAVV	Rapport des coefficients d’ondelettes [15]	WAVR
Variance de niveau de gris [4]	GLVA	Variance normalisée du niveau de gris [11]	GLVN
Image contrast [6]	CONT		

TABLE 1 – Abréviations des opérateurs de mesure de focalisation utilisés dans les expériences.

### 2.1 Results

Le tableau 2 rapporte les résultats obtenus sur les ensembles de validation et de test. Pour l’approche DFF, il n’y a pas d’opérateur qui conserve ses performances pour les différentes scènes ; les performances des opérateurs vont en effet dépendre de la géométrie de la scène. Pour les scènes avec occultations, certains objets peuvent être positionnés l’un derrière l’autre, conduisant à plusieurs niveaux de focalisation sur les patches extraits aux bords de l’objet ; dans ce cas, la prédiction d’une profondeur unique pour le pixel central donne de mauvaises performances. Le rapport entre

	Piano	Table	Woods	Dices	Cars
<b>Approche basée sur les patches</b>					
U-Net	<b>0.66</b>	<b>1.12</b>	<b>0.73</b>	<b>1.21</b>	<b>3.48</b>
<b>Opérateurs basés sur le laplacien</b>					
LAPE	11.58	25.81	16.35	<b>5.83</b>	32.56
LAPM	18.89	16.48	23.158	27.0	51.26
LAPV	<b>4.12</b>	<b>9.28</b>	<b>3.82</b>	<b>2.84</b>	<b>17.07</b>
LAPD	23.88	19.93	26.81	26.81	54.16
<b>Opérateurs basés sur les ondelettes</b>					
WAVV	9.95	11.24	<b>7.19</b>	13.26	23.88
WAVR	8.16	8.98	7.45	8.76	<b>14.44</b>
<b>Opérateurs divers</b>					
GLVN	17.90	15.10	20.70	25.75	51.65
GLVA	21.40	17.10	19.38	24.12	51.62
CONT	16.33	14.89	19.51	21.30	47.13

TABLE 2 – Norm  $\ell_1$  en utilisant différentes approches pour les ensembles de validation et de test.

les patches texturés et non texturés joue également un rôle sur la performance finale, certains opérateurs seront plus performants sur les zones fortement texturées mais échoueront sur les zones uniformes. Globalement, la variance laplacienne et les approches basées sur les ondelettes sont les opérateurs les plus performants.

L’approche proposée basée sur les patches est plus précise et plus rapide que la méthode DFF classique car elle peut utiliser une seule inférence par lot pour déterminer la carte de profondeur pour chaque découpe du volume de reconstruction. En outre, elle est plus précise dans la segmentation des différentes sous-régions du volume de reconstruction recadré, et ne nécessite pas de post-traitement supplémentaire ou de taille de patch adaptative par pixel pour maintenir la continuité dans la carte de profondeur finale. Le réseau est entraîné à identifier les caractéristiques pertinentes d’une image focalisée en utilisant des patches d’entrée de taille fixe et à la segmenter correctement. En outre, le réseau garantit que le niveau de focalisation de chaque pixel suit une distribution gaussienne parfaite, avec des pics correspondant à la focalisation optimale, et n’est donc pas sujet à des changements de polarité lorsque la taille du patch ou la plage de reconstruction est trop grande.

## 3 Conclusion

Dans cet article, nous proposons une approche basée sur l’apprentissage profond pour l’estimation de la profondeur à partir d’hologrammes générés par ordinateur et nous comparons sa performance aux méthodes classiques de DFF.

L’approche donne les résultats de prédiction de profondeur les plus précis. Elle permet de segmenter uniquement les parties qui sont focalisées sur chaque reconstruction holographique, ce qui permet un traitement de mise au point à plusieurs niveaux dans le même patch, tout en maintenant des bords et des contours cohérents entre les différentes régions extraites.

## Références

- [1] David Blinder, Ayyoub Ahar, Stijn Bettens, Tobias Birnbaum, Athanasia Symeonidou, Heidi Ottevaere, Colas Schretter, and Peter Schelkens. Signal processing challenges for digital holographic video display systems. *Signal Processing : Image Communication*, 70 :114–130, 2019.
- [2] Antonin Gilles, Patrick Gioia, Rémi Cozot, and Luce Morin. Hybrid approach for fast occlusion processing in computer-generated hologram calculation. *Appl. Opt.*, 55(20) :5459–5470, Jul 2016.
- [3] Joseph W Goodman. Introduction to fourier optics. *Introduction to Fourier optics, 3rd ed., by JW Goodman. Englewood, CO : Roberts & Co. Publishers, 2005*, 1, 2005.
- [4] Eric Krotkov and Jean-Paul Martin. Range from focus. *Proceedings. 1986 IEEE International Conference on Robotics and Automation*, 3 :1093–1098, 1986.
- [5] Andreas Erik Gejl Madsen, René Lynge Eriksen, and Jesper Glückstad. Comparison of state-of-the-art computer generated holography algorithms and a machine learning approach. *Optics Communications*, 505 :127590, 2022.
- [6] Harsh Nanda and Ross Cutler. Practical calibrations for a real-time digital omnidirectional camera. *Proceedings of CVPR, Technical Sketch*, 01 2001.
- [7] S.K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8) :824–831, 1994.
- [8] J.L. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia. Diatom autofocusing in brightfield microscopy : a comparative study. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 3, pages 314–317 vol.3, 2000.
- [9] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5) :1415–1432, 2013.
- [10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
- [11] A Santos, Carlos Ortiz-de Solorzano, Juan Jose Vaquero, J Peña, Norberto Malpica, and Francisco Del Pozo Guerrero. Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of microscopy*, 188 :264–72, 01 1998.
- [12] Murali Subbarao, Tae Choi, and Arman Nikzad. Focusing techniques. *Journal of Optical Engineering*, 32 :2824–2836, 1993.
- [13] Athanasia Symeonidou, David Blinder, Ayyoub Ahar, Colas Schretter, Adrian Munteanu, and Peter Schelkens. Speckle noise reduction for computer generated holograms of objects with diffuse surfaces. In Peter Schelkens, Touradj Ebrahimi, Gabriel Cristóbal, Frédéric Truchetet, and Pasi Saarikko, editors, *Optics, Photonics and Digital Technologies for Imaging Applications IV*, volume 9896, page 98960F. International Society for Optics and Photonics, SPIE, 2016.
- [14] Andrea Thelen, Susanne Frey, Sven Hirsch, and Peter Hering. Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation. *IEEE Transactions on Image Processing*, 18(1) :151–157, 2009.
- [15] Ge Yang and B.J. Nelson. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, volume 3, pages 2143–2148 vol.3, 2003.
- [16] Yan Zhao, Liangcai Cao, Hao Zhang, Dezhao Kong, and Guofan Jin. Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method. *Opt. Express*, 23(20) :25440–25449, Oct 2015.