

# Reconstruction 3D depuis des couples SAR ascendant/descendant

Emile BARBIER--RENARD<sup>1</sup> Florence TUPIN<sup>1</sup> Loïc DENIS<sup>2</sup>

<sup>1</sup>LTCl, Télécom Paris, 91120 Palaiseau, France

<sup>2</sup>Laboratoire Hubert Curien, UMR 5516, CNRS, UJM, IOGS, 42000 Saint-Étienne, France

**Résumé** – Afin d’exploiter les disparités géométriques entre des images SAR acquises sous différents points de vue, nous présentons une méthode de reconstruction de surface par rendu inverse inspirée de *NeRF*. Elle optimise une carte d’élévation et une carte de coefficients de rétrodiffusion à partir d’un minimum de deux images, et s’appuie sur un modèle de rendu différentiable adapté à cette représentation en carte d’élévation ainsi qu’une stratégie multi-échelles assurant une convergence rapide. Nous validons les capacités de reconstruction sur des données synthétiques réalistes générées par le simulateur EMPRISE<sup>®</sup> de l’ONERA.

**Abstract** – To exploit the geometric disparities between SAR images acquired from multiple points of view, we introduce a surface reconstruction by inverse rendering method inspired by *NeRF*. It optimises a DSM and a map of backscattering coefficients from as few as two images, and relies on a custom differentiable SAR rendering model as well as a coarse-to-fine strategy ensuring fast convergence. We verify the reconstruction abilities on realistic synthetic data generated by ONERA’s EMPRISE<sup>®</sup> simulator.

## 1 Introduction

### 1.1 Reconstruction par rendu inverse

La reconstruction 3D par rendu inverse est un développement récent popularisé par les *Neural Radiance Fields (NeRF)* [9], permettant l’extraction de la géométrie et de l’apparence d’une scène depuis un ensemble arbitraire de vues. Celle-ci repose sur un modèle de rendu physique mettant en lien une géométrie à optimiser et les observations disponibles : par une méthode de descente de gradient stochastique on minimise l’écart entre les images observées et des vues synthétiques calculées sur la représentation de la scène. Cette approche prodigue une grande flexibilité dans les vues exploitées et permet des reconstructions à partir d’un nombre limité d’images. Afin de pouvoir être employé dans le cadre d’une reconstruction par rendu inverse, le modèle de rendu utilisé doit être :

- différentiable pour permettre son inversion ;
- rapide à calculer pour effectuer beaucoup d’itérations.

Dans sa formulation initiale, *NeRF* [9] représente la scène par un champ de densité 3D (où les zones de haute densité correspondent aux objets solides) et un champ de radiance (caractérisant la couleur en chaque point de l’espace, en prenant en compte l’angle d’observation). Pour assurer une représentation continue, ces fonctions multidimensionnelles sont modélisées par un perceptron multicouches dont les paramètres constituent les poids à optimiser. Les images sont calculées avec un algorithme de rendu volumique intégrant la contribution de nombreux points échantillonnés le long d’un rayon.

Suite à son introduction en 2020, de nombreux travaux ont poursuivi la piste ouverte par *NeRF* afin d’améliorer ses performances et vitesses de reconstruction, ou de l’adapter à des

usages spécifiques. En télédétection, les études ont principalement porté sur l’extension aux images satellitaires optiques : les travaux de Marí *et al.* [7, 8] mettent en avant la complexité liée à la variabilité de l’apparence des images causée par la position du soleil ou la météo. En plus de la densité et d’une albédo communes, ils suggèrent donc d’optimiser des termes spécifiques à chaque image expliquant ces phénomènes.

Il semble donc intéressant d’adapter cette approche de reconstruction par rendu inverse à l’imagerie radar à synthèse d’ouverture (SAR). En effet, contrairement à l’imagerie optique, l’imagerie SAR est un procédé actif produisant sa propre illumination ; elle est ainsi indépendante du soleil et traverse les couches nuageuses. De plus les images SAR sont riches en informations géométriques : des phénomènes d’ombre et de repliement. Réciproquement, la reconstruction par rendu inverse présente des avantages sur les méthodes traditionnelles de reconstruction en imagerie SAR.

### 1.2 Reconstruction de surface en imagerie SAR

Les principales approches multi-images comme l’interférométrie [1] ou la tomographie [4] requièrent des acquisitions cohérentes issues de vues très similaires. Les approches incohérentes, exploitant les intensités des échos, appartiennent des points isolés associés à des éléments comme des points brillants ([10]), ou reconstruisent une surface depuis une unique image en supposant la scène homogène (radarclinométrie [11])

La reconstruction par rendu inverse appliquée au SAR a donc été peu étudiée jusqu’à très récemment. Deux articles ont proposé en 2024 une adaptation du rendu volumique de *NeRF* : pour simuler une ligne radar, un ensemble de rayons est lancé et la scène est échantillonnée le long de chaque rayon en des positions correspondant aux différentes cases distances de l’image SAR. L’intensité de l’image simulée est obtenue par l’accumulation de la contribution des éléments de volume échantillonnés le long de chaque rayon. *SARNeRF* [5] optimise une grille voxel 3D de densité et d’intensité diffusée, mais s’intéresse plus à la synthèse de nouvelles vues qu’à la reconstruction de surface. Dans *Radar Fields*, Ehret *et al.* [3] propose

Ce travail est financé par l’Agence de l’Innovation de Défense (AID) et par le projet Apprentissage statistique pour l’imagerie SAR multidimensionnelle (ASTRAL) sous la bourse ANR-21-ASTR-0011. Nous remercions également l’ONERA pour l’accès au simulateur EMPRISE<sup>®</sup> : <https://www.emprise-em.fr>

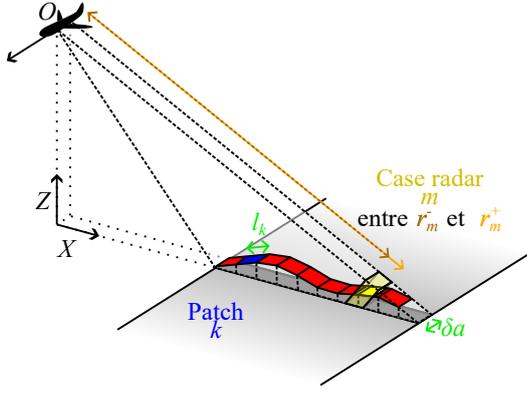


FIGURE 1 : Géométrie du rendu proposé. La surface est discrétisée en patchs rectangulaires  $k$  inscrits dans le faisceau radar, dont on estime la contribution à chaque case radar  $m$ .

l'apprentissage de cartes d'élévation et de coefficients de spécularité. Afin d'appliquer un rendu volumique, une densité est estimée depuis la carte d'élévation. Les résultats présentés sont prometteurs mais ont à ce jour été validés uniquement sur des images issues du propre modèle de rendu.

### 1.3 Contributions

Nous présentons une méthode de reconstruction de surface par rendu inverse adaptée aux images SAR en intensité. Afin d'estimer un modèle numérique de surface (MNS), nous représentons la scène par une carte d'élévation et une carte de coefficients de rétrodiffusion. Nous proposons un modèle de rendu d'images SAR différentiable reposant sur un échantillonnage direct de la surface, en opposition à l'échantillonnage à travers le volume nécessaire pour les modèles issus du rendu volumique. Nous modélisons chacune des cartes par une grille multi-échelles afin d'appliquer une stratégie « grossier-vers-fin » à notre optimisation par descente de gradient stochastique. Nous présentons une méthode de reconstruction de surface :

1. s'appuyant uniquement sur l'intensité des images SAR ;
2. applicable même à un nombre très limité de vues ;
3. exploitant un modèle de rendu différentiable en une seule passe, échantillonnant directement la surface ;
4. employant une stratégie multi-échelles limitant les artefacts et non biaisée par une régularisation ;
5. sans pré-entraînement sur des données externes.

Nous démontrons ses capacités à extraire un MNS depuis seulement 2 images en configuration ascendant/descendant par des reconstructions sur des données synthétiques réalistes produites par le simulateur EMPRISE® de l'ONERA [13].

## 2 Modèle de rendu SAR différentiable

### 2.1 Formulation

Considérons une acquisition SAR Stripmap présentant une réponse surfacique (typiquement en bande X ou Ka où la pénétration est négligeable), permettant de représenter la géométrie de la scène sous la forme d'un MNS  $\mathcal{Z}(x, y)$  et son apparence par une carte de coefficients de rétrodiffusion  $\mathcal{B}(x, y)$ .

Nous proposons de discrétiser la surface couverte par le faisceau radar synthétique (après synthèse d'ouverture) en  $K$  patchs rectangulaires tels qu'illustré figure 1, et de calculer leurs contributions à chacune des cases radar. Nous les formons par échantillonnage uniforme au sol de  $K + 1$  points  $P_k$  dans le plan zéro-Doppler et évaluation de  $\mathcal{Z}$  pour trouver leur élévation puis nous joignons le point  $P_k$  au point  $P_{k+1}$ . Les segments de longueur  $l_k$  obtenus sont étirés dans la direction azimutale de la largeur du faisceau radar  $\delta a$ . Nous formulons ainsi l'espérance de l'intensité de la  $m$ ème case radar par :

$$\hat{I}_m = \sum_{k=0}^{K-1} \sigma_k^0 \cdot \hat{v}_k \cdot \hat{w}_{m,k} \cdot (l_k \cdot \delta a). \quad (1)$$

La contribution de chaque patch d'aire  $l_k \cdot \delta a$  dépend de :

- $\sigma_k^0$ , sa surface équivalente radar normalisée, pour un matériau supposé lambertien. Elle correspond donc à  $\mathcal{B}$  évalué en son centre et pondéré par une atténuation en cosinus entre la normale du patch et la ligne de visée ;
- $\hat{v}_k$ , son indicatrice d'éclairage, tendant vers 0 s'il est situé dans l'ombre ;
- $\hat{w}_{m,k}$ , l'indicatrice de la  $m$ ème case radar, non négligeable pour les segments situés à la distance adéquate.

### 2.2 Prise en compte des ombres

Appelons  $\hat{v}_k$  l'indicatrice d'éclairage associée au patch formé par les points échantillonnés en  $P_k$  et  $P_{k+1}$ . Pour simplifier l'évaluation, nous considérons le patch éclairé si son extrémité lointaine  $P_{k+1}$  est elle-même éclairée.

Il est possible de calculer en une seule passe l'éclairage de tous les points en les parcourant par distance au sol croissante : en effet un point  $P_{k+1}$  est éclairé s'il n'est pas à l'ombre de  $P_j$ ,  $j \leq k$ , le dernier point éclairé. Géométriquement, cela revient à vérifier si  $P_{k+1}$  est situé au dessus de la droite de démarcation des ombres  $(0, P_j)$  liant le capteur au point  $P_j$ . Nous effectuons ce calcul en comparant la pente  $\eta_{k+1}$  de la droite  $(O, P_{k+1})$  à  $h_k = \eta_j$  la pente la ligne de démarcation des ombres puis en actualisant  $h_{k+1}$ , en choisissant  $h_0 = \eta_0$  :

$$\begin{cases} \hat{v}_k = \mathcal{S}(\eta_{k+1} - h_k) \\ h_{k+1} = \eta_{k+1} \cdot \hat{v}_k + h_k \cdot (1 - \hat{v}_k), \end{cases} \quad (2)$$

où  $\mathcal{S}$  approxime un test binaire par une sigmoïde logistique.

### 2.3 Contribution aux cases radar

Une case radar  $m$  accumule les échos étant réfléchis à des distances comprises dans l'intervalle  $R_m = [r_m^-, r_m^+]$ . Considérons le patch  $k$  formé depuis les points  $P_k$  et  $P_{k+1}$  de distances au capteur respectives  $d_k^-$  et  $d_k^+$ . On estime la contribution  $\hat{w}_{m,k}$  du patch  $k$  à la case  $m$  en projetant le patch sur la ligne de visée : elle correspond alors à la proportion de  $[d_k^-, d_k^+]$  intersectant  $R_m$ . On évalue cette contribution en utilisant  $\mathcal{M}$  une approximation différentiable du maximum :

$$\hat{w}_{m,k} = \frac{\mathcal{M}(d_k^-, r_m^+) + \mathcal{M}(d_k^+, r_m^-)}{d_k^+ - d_k^-} - \frac{\mathcal{M}(d_k^+, r_m^+) + \mathcal{M}(d_k^-, r_m^-)}{d_k^+ - d_k^-}. \quad (3)$$

### 3 Reconstruction de surface

#### 3.1 Modélisation de la scène

Nous modélisons chacune des cartes à optimiser (élévation et rétrodiffusion) par une grille multi-échelles, analogue à la décomposition en ondelettes d'une image. Elle est constituée de  $L$  matrices de paramètres  $G_l$ ,  $l \in \llbracket 1, L \rrbracket$  de dimensions respectives  $2^l \times 2^l$ , interpolées bilinéairement par  $\mathcal{G}_l(x, y)$ . L'évaluation de la carte en un point  $(x, y)$  correspond à :

$$\mathcal{C}_s(x, y) = \sum_{l=1}^L \mathcal{G}_l(x, y) \cdot \frac{1}{2^l} \cdot w_s(l). \quad (4)$$

Afin de favoriser initialement la reconstruction des basses-fréquences, nous choisissons de ne pas employer de terme de régularisation (comme une variation totale) biaisant notre optimisation. Nous retardons plutôt la contribution des hautes fréquences en pondérant chaque échelle par  $\frac{1}{2^l}$ , permettant ainsi d'estimer d'abord les basses-fréquences. Ce phénomène est renforcé par la deuxième pondération  $w_s(l)$  correspondant à un masquage des fréquences au delà d'une échelle  $s$  choisie, telle que définie par *Nerfies* [12, 6] (alors appliquée aux bandes de fréquence issues d'un encodage positionnel) :

$$w_s(l) = \frac{1 - \cos(\pi \cdot \min(\max(s - l, 0), 1))}{2}. \quad (5)$$

Les échelles plus grossières que  $\lfloor s \rfloor$  contribuent pleinement tandis que celles plus fines que  $\lceil s \rceil$  sont omises. L'échelle seuil  $s$  correspond à  $s_d + s_b$  où :

- $s_d$  est l'échelle déduite de l'écart entre deux échantillons lors du rendu, déterminant les fréquences utiles ;
- $s_b$  croît linéairement au cours de l'optimisation, introduisant progressivement des hautes fréquences.

Une activation exponentielle est appliquée à la carte de rétrodiffusion pour garantir des résultats positifs.

#### 3.2 Protocole

Pour nos expériences, effectuées sur une carte RTX 3080, nous avons appliqué 400 itérations de l'algorithme d'optimisation Adam. Pour chacune, nous effectuons le rendu de  $M = 86$  lignes de  $N$  cases radar choisies aléatoirement parmi toutes les images. Le taux d'apprentissage décroît linéairement sur  $[2 \cdot 10^{-3}, 2 \cdot 10^{-2}]$ , et  $s_b$  croît sur  $[-4, 4]$ . Les grilles multi-échelles ont un niveau maximum  $L = 11$  ( $2048 \times 2048$ px).

La fonction de coût correspond à l'opposé de la log-vraisemblance entre les intensités observées  $I$  et prédites  $\hat{I}$ , prenant en compte la distribution du chatoiement [2], est :

$$\mathcal{L}(\hat{I}, I) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \log \left( \frac{\hat{I}_{m,n}}{I_{m,n}} \right) + \frac{I_{m,n}}{\hat{I}_{m,n}}. \quad (6)$$

#### 3.3 Résultats

Afin de mettre en lumière les capacités de reconstruction de notre méthode, nous avons appliqué notre approche sur un ensemble d'images générées par le simulateur radar EMPRISE<sup>®</sup> de l'ONERA, d'angle d'incidence en milieu de fauchée de  $45^\circ$  et de lignes de 1002 cases radar mesurant 1,5m en azimuth

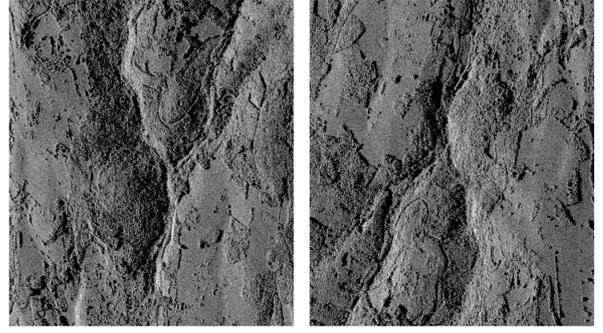


FIGURE 2 : Vues ascendant/descendant, générées avec EMPRISE<sup>®</sup>, exploitées pour la reconstruction à deux images. L'axe des distances est horizontal (capteur à gauche).

et en distance. Ces vues correspondent aux Gorges du Désert de Saint-Alban-les-Eaux, France, à partir du MNS produit par l'IGN-F dans le cadre de la mission LiDAR-HD<sup>1</sup>.

Une première reconstruction, à partir uniquement des deux vues présentées figures 2, met en avant le potentiel de l'approche pour l'exploitation d'acquisitions en configuration ascendant/descendant. Les résultats à différentes phases de l'optimisation sont présentés figure 3 et traduisent une convergence rapide en moins de 10 minutes. Le MNS retrouvé présente une erreur RMS de 5,55m sur une plage de hauteurs de plus de 250m, vérifiant la fidélité visuelle de la reconstruction.

L'évolution du terrain au fil des itérations confirme la reconstruction initiale des basses-fréquences telle que recherchée avec notre approche grossier-vers-fin. Si comme attendu les bords ne sont pas restitués avec une extrême netteté, notre méthode parvient toutefois à retrouver des structures fines telles que la route longeant le relief sud-ouest.

Une seconde reconstruction de la même scène, effectuée sur 5 images acquises selon une trajectoire circulaire, atteint une RMSE de 3,82m, sans altération des hyper-paramètres ou du temps de calcul. Cela caractérise le bénéfice que tire cette approche de la diversité de points de vue.

### 4 Conclusion

Dans cet article nous avons présenté sur des données simulées réalistes EMPRISE<sup>®</sup> une approche de reconstruction 3D en imagerie SAR reposant sur le rendu inverse, et montré sa qualité et vitesse de reconstruction même dans le cas d'un nombre très limité d'images. Outre l'avantage apporté dans la reconstruction de structures d'échelles variées, la modélisation des cartes d'élévation et de rétrodiffusion par des grilles multi-échelles offre également une bonne interprétabilité des paramètres optimisés en comparaison aux modélisations par réseau de neurone. Suite à ces résultats encourageants, l'application sur des observations SAR réelles constitue la prochaine phase de recherche notamment en assurant une résilience face à des phénomènes tels que la présence de lobes secondaires, de doubles rebonds ou de réponses directionnelles. La flexibilité apportée par l'approche développée ouvre également des perspectives d'exploitation conjointe de capteurs SAR différents.

### Références

- [1] Richard BAMLER et Philipp HARTL : Synthetic Aperture Radar interferometry. *Inverse Problems*, 14(4):R1, 1998.

<sup>1</sup>IGN-F : "MNS LiDAR HD," 2024.

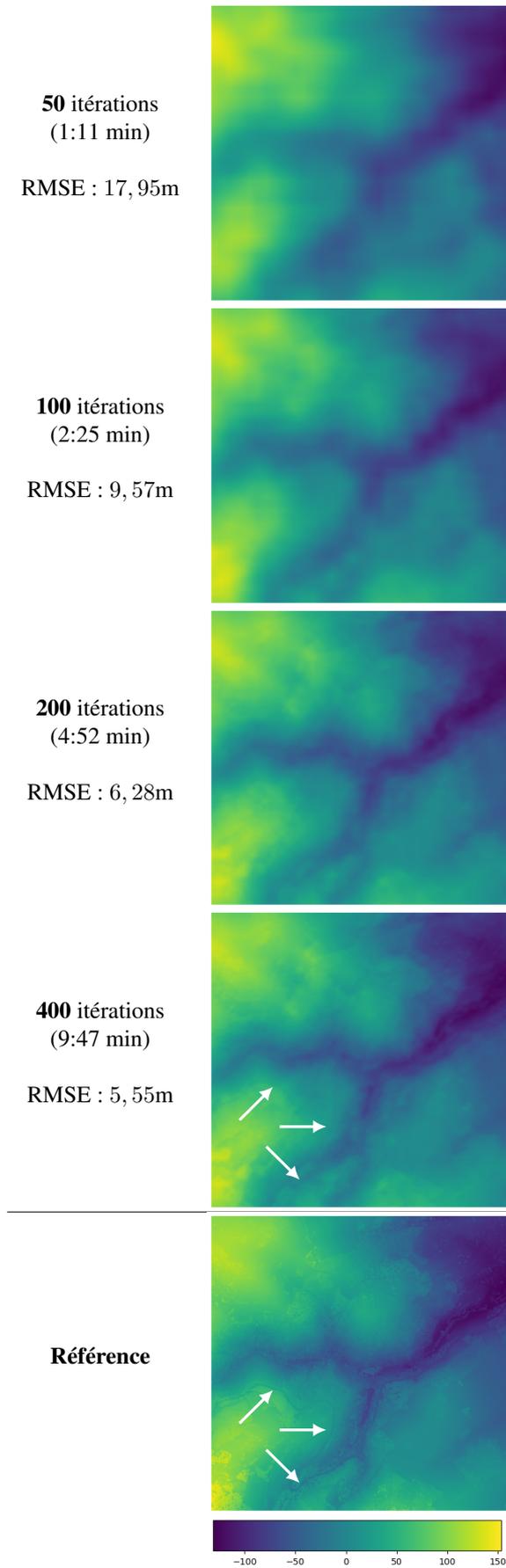


FIGURE 3 : Résultats de la reconstruction de la Gorge du Désert à partir de 2 images SAR synthétiques à plusieurs phases de l'optimisation. L'itération est indiquée au dessus, la RMSE en dessous. Les flèches indiquent le tracé d'une route.

- [2] Emanuele DALSASSO, Loïc DENIS et Florence TUPIN : SAR2SAR : A semi-supervised despeckling algorithm for SAR images. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 14:4321–4329, 2021.
- [3] Thibaud EHRET, Roger MARI, Dawa DERKSEN, Nicolas GASNIER et Gabriele FACCILOLO : Radar Fields : An Extension of Radiance Fields to SAR. *In IEEE/CVF CVPR Workshops*, pages 564–574, 2024.
- [4] Gianfranco FORNARO, Fabrizio LOMBARDINI, Antonio PAUCIULLO, Diego REALE et Federico VIVIANI : Tomographic processing of interferometric SAR data : Developments, applications, and future research perspectives. *IEEE Signal Process. Mag.*, 31(4):41–50, 2014.
- [5] Zhengxin LEI, Feng XU, Jiangtao WEI, Feng CAI, Feng WANG et Ya-Qiu JIN : SAR-NeRF : Neural Radiance Fields for Synthetic Aperture Radar Multiview Representation. *IEEE Trans. Geosci. Remote Sens.*, 62:1–15, 2024.
- [6] Chen-Hsuan LIN, Wei-Chiu MA, Antonio TORRALBA et Simon LUCEY : BARF : Bundle-Adjusting Neural Radiance Fields. *In IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 5721–5731, 2021.
- [7] Roger MARÍ, Gabriele FACCILOLO et Thibaud EHRET : Sat-NeRF : Learning multi-view satellite photogrammetry with transient objects and shadow modeling using RPC cameras. *In IEEE/CVF CVPR Workshops*, pages 1310–1320, 2022.
- [8] Roger MARÍ, Gabriele FACCILOLO et Thibaud EHRET : Multi-date Earth observation NeRF : The detail is in the shadows. *In IEEE/CVF CVPR Workshops*, pages 2034–2044, 2023.
- [9] Ben MILDENHALL, Pratul P. SRINIVASAN, Matthew TANCIK, Jonathan T. BARRON, Ravi RAMAMOORTHY et Ren NG : NeRF : Representing scenes as neural radiance fields for view synthesis. *In ECCV*, 2020.
- [10] Stephan PALM, Hélène M. ORIOU et Hubert M. CANTALOUUBE : Radargrammetric DEM extraction over urban area using circular SAR imagery. *IEEE Trans. Geosci. Remote Sens.*, 50(11):4720–4725, 2012.
- [11] S. PAQUERAULT, H. MAITRE et J.-M. NICOLAS : Radarclinometry for ERS-1 data mapping. *In IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, volume 1, pages 503–505 vol.1, 1996.
- [12] Keunhong PARK, Utkarsh SINHA, Jonathan T. BARRON, Sofien BOUAZIZ, Dan B GOLDMAN, Steven M. SEITZ et Ricardo MARTIN-BRUALLA : Nerfies : Deformable neural radiance fields. *In IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, pages 5845–5854, 2021.
- [13] Nicolas TROUVÉ, Christian COCHIN, Julien HOUSSAY, Etienne EVERAERE, Xavier HUSSON, Kevin UNGERS, Antoine JOUADÉ, Ronan FABBRI, Thomas HOURET, Benoit TALIBART, Olivier LÉVÊQUE et Xavier DUPUIS : EMPRISE : Synthetic Environment for sensor design and virtual qualification. *In IEEE Int. Radar Conf. (RADAR)*, 2024.