



Université
Gustave Eiffel

THALES

Apprentissage profond de la similarité pour la reconstruction 3D large échelle: Cas des images satellitaires Pléaides Néo

Mohamed Ali Chebbi^{1,2}

Supervisé par:

Marc Pierrot-Deseilligny¹, Ewelina Rupnik¹ et Paul Lopes²

¹ LaSTIG, UGE, ENSG, IGN

² Thales

Problématique

Objectifs

Démarche méthodologique

Généralisation: Aérien vers PNEO

Adaptation de domaine par fine-tuning

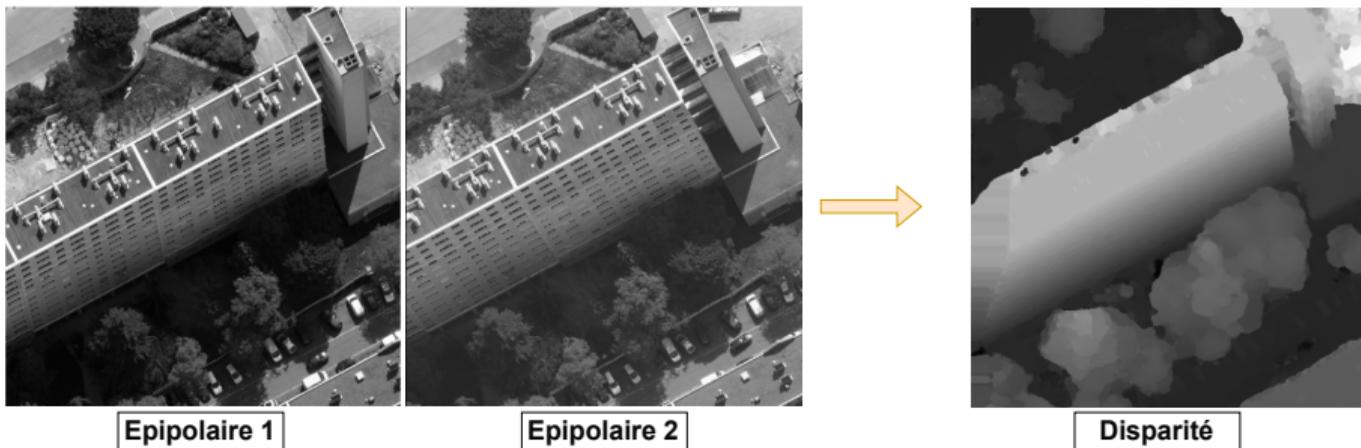
Perspectives



Problématique

Appariement dense/Mise en correspondance (MEC) ?

- ▶ Processus de mise en correspondance **pixel à pixel** entre **2 ou plusieurs** images[1].



Appariement en géométrie épipolaire.

2 étapes fondamentales

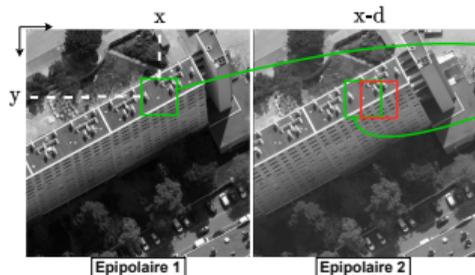
- ▶ Calcul du cube de coût/corrélation: se base sur une métrique de corrélation entre des vignettes d'images.
 - ▶ SSD, NCC, Census, **Similarité apprise** [2].
- ▶ Régularisation du cube de corrélation par des méthodes semi-globales [1, 3] ou globales [4].

Apprentissage

Apprentissage end-to-end de la disparité

Apprentissage de similarité

1. Mesure de similarité sur des vignettes d'images: 3x3, 5x5, 11x11



Fonction de similarité

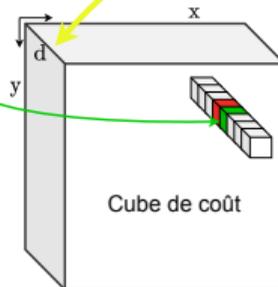
Classique:

- SSD
- NCC
- census

Apprentissage:

- CNN

Intervalle de recherche



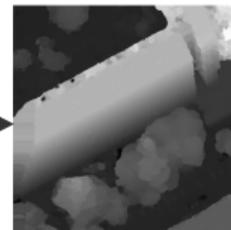
$$E(D) = \sum_p C(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1]$$

Attache aux données

Apprentissage de régularisation

2. Cohérence spatiale des correspondances imposées par la régularisation

Régularisation



Disparité

Termes de régularité

$\sum_{q \in N_p} P_1 T[|D_p - D_q| = 1]$ Pénalisation des surfaces inclinées
 $\sum_{q \in N_p} P_2 T[|D_p - D_q| > 1]$ Pénalisation des sauts de disparité

N_p est le voisinage d'un pixel p , D_p et D_q sont les disparités en p and q , respectivement. On pénalise les surfaces en rampe à P_1 . Les sauts de disparité > 1 sont pénalisés à P_2 .

Objectifs

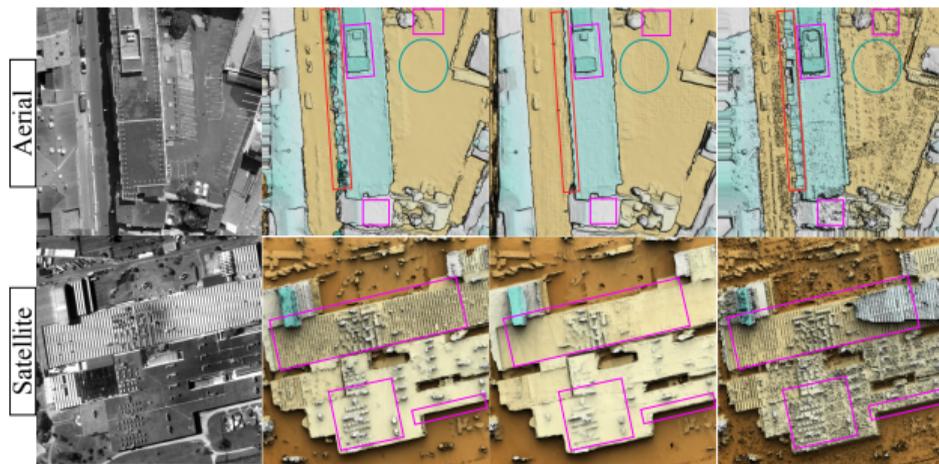
Objectif I: Apprendre des similarités **denses** et **consistantes** sur l'**ensemble** d'un support de l'image.

Objectif II: Résoudre le problème d'appariement majoritairement dans la phase de calcul de similarité.

Objectif III: Adapter l'apprentissage aux différentes géométries d'appariement (géométrie faisceau).

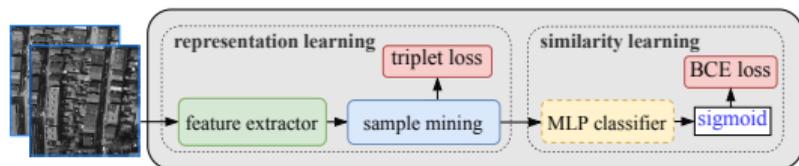
Démarche méthodologique

- ▶ DeepSim-Nets est capable de séparer **sans ambiguïté** les pixels homologues des pixels non homologues.
- ▶ Projet: <https://dalichebbi.github.io/DeepSimNets/>



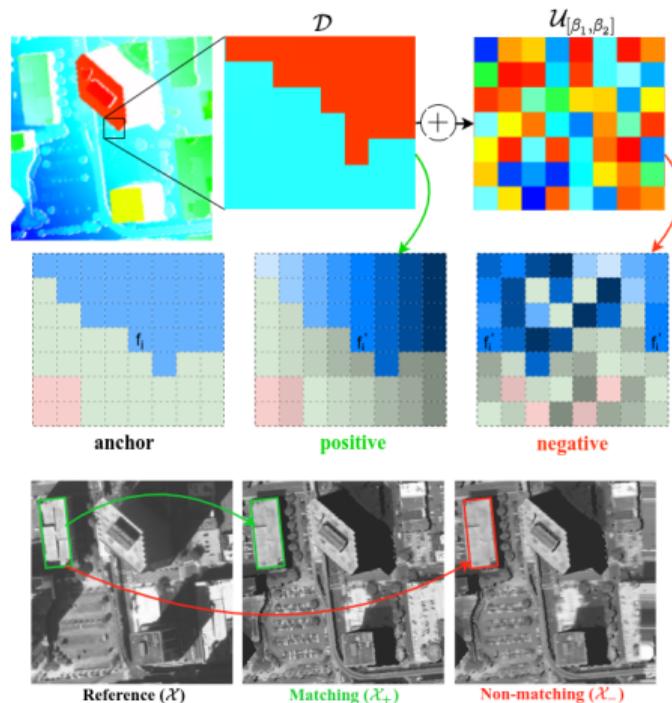
((a)) Épip.1 ((b)) **OURS** ((c)) PSMNet ((d)) NCC

Demo.



Pipeline proposé.

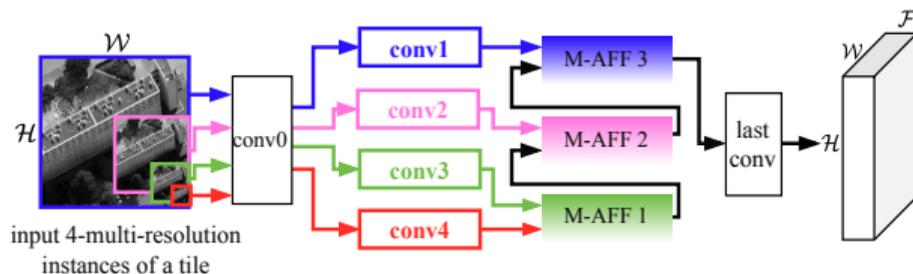
- ▶ Un CNN profond pour calculer les descripteurs: U-Net32, U-Net Attention, MS-AFF (proposé).
- ▶ Affectation de l'ensemble des descripteurs de l'épipoilaire 1 à leurs correspondants dans l'épipoilaire 2.
- ▶ Apprentissage contrastif.



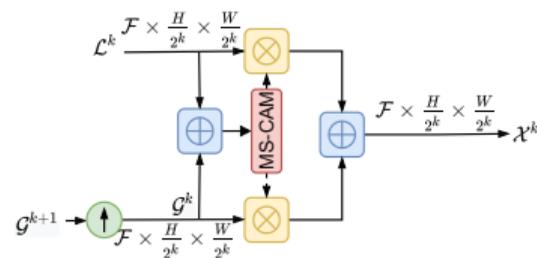
Génération d'exemples denses pour l'entraînement des réseaux à large champ récepteur.

Extracteurs de représentations

- ▶ U-Net 32
- ▶ U-Net Attention
- ▶ MS-AFF : backbone proposé utilisant un mécanisme d'attention multi-échelle.



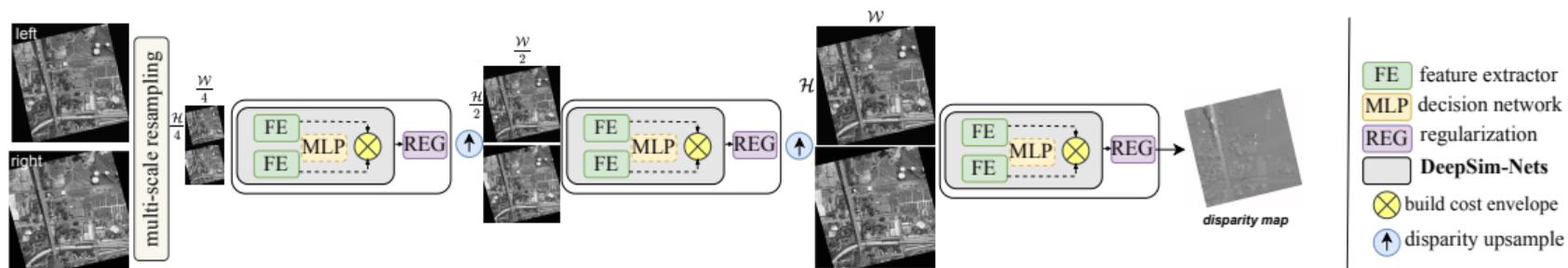
Architecture MS-AFF.



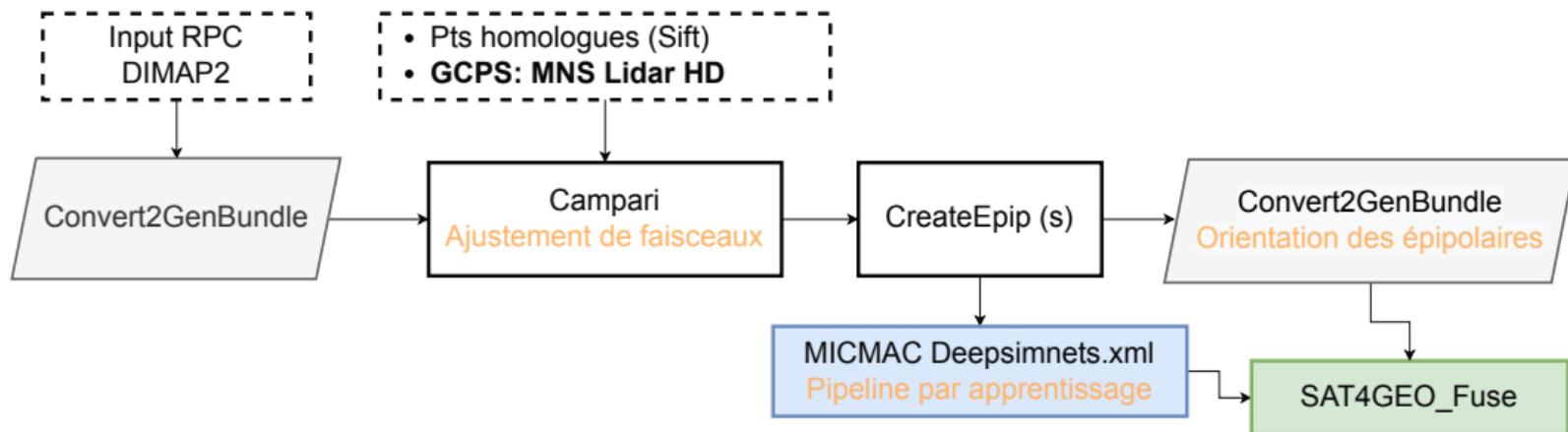
Module de fusion entre deux échelles consécutives (M-AFF).

Pipeline multi-échelle hybride

- ▶ NCC pour des faibles résolutions: robustesse, bon a priori.
- ▶ Similarités apprises à partir de la résolution $\frac{1}{4}$: granularité et distinctivité dans les zones homogènes.



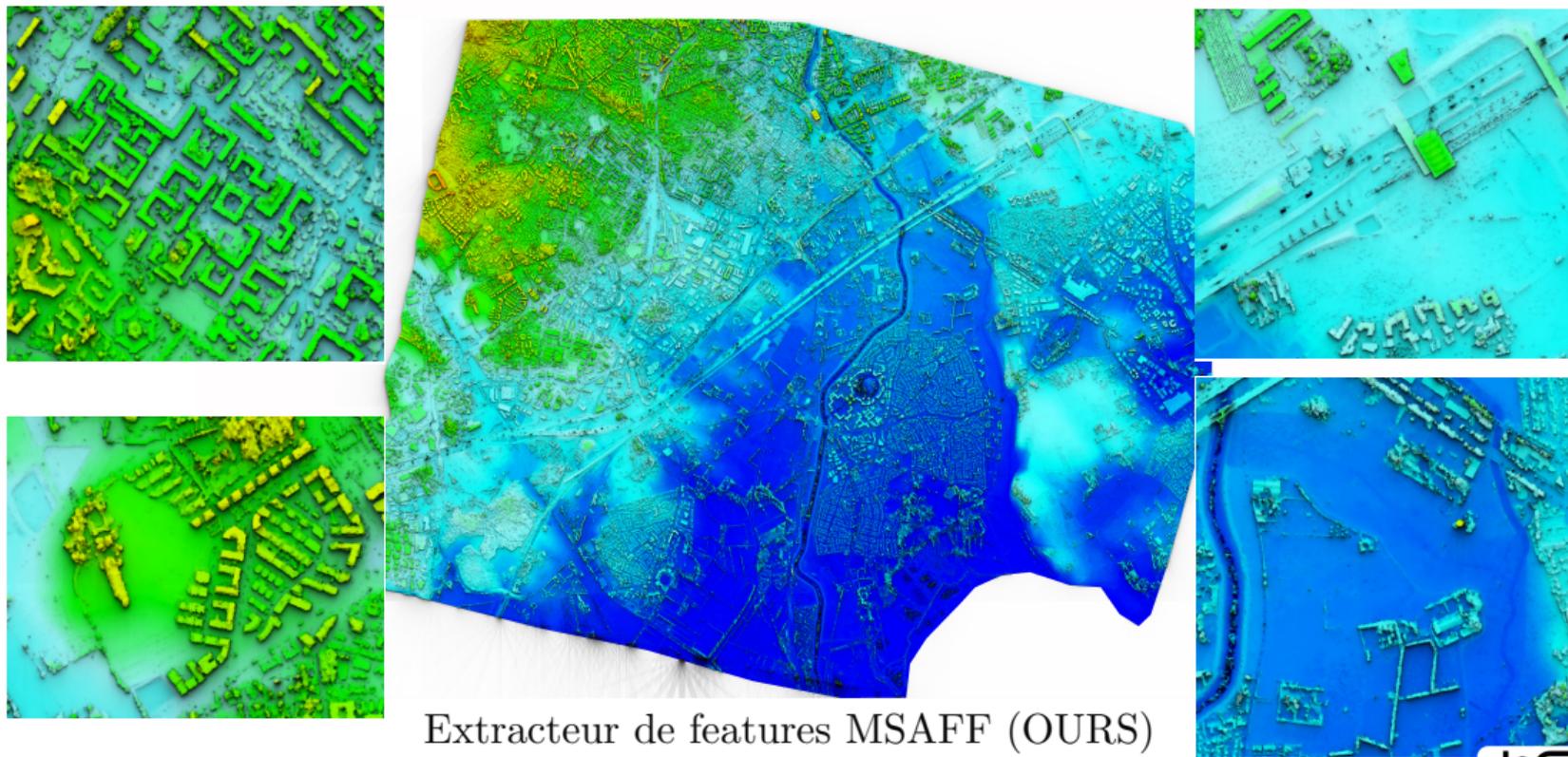
Intégration des modèles de similarité dans le pipeline multi-échelle de MicMac.



Pipeline fonctionnel dans MicMac: Intégration de DeepSim-Nets.

Généralisation: Aérien vers PNEO

- ▶ Reconstruction de MNS PNEO:
 - ▶ Cas d'étude: Tri-Stéréo PNEO sur Montpellier (GSD=0.32 m).
 - ▶ Approche de calcul par **fusion** après passage à la géométrie **épipolaire**.
 - ▶ 2 cartes de disparité seulement par triplet.



Extracteur de features MSAFF (OURS)

Qualité des mesures de similarité apprises



DeepSim-Nets

Similarité apprise vs. corrélation classique

Comportement à l'échelle unitaire d'une carte de disparité.

Image de référence

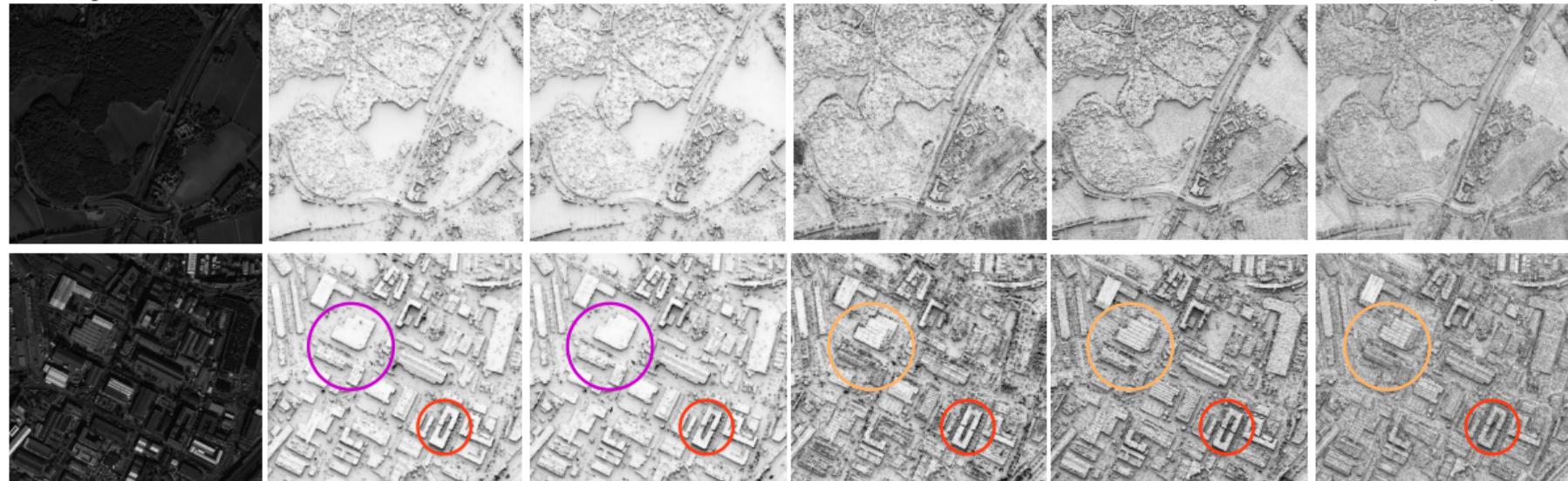
UNet

UNet Attention

MCCNN

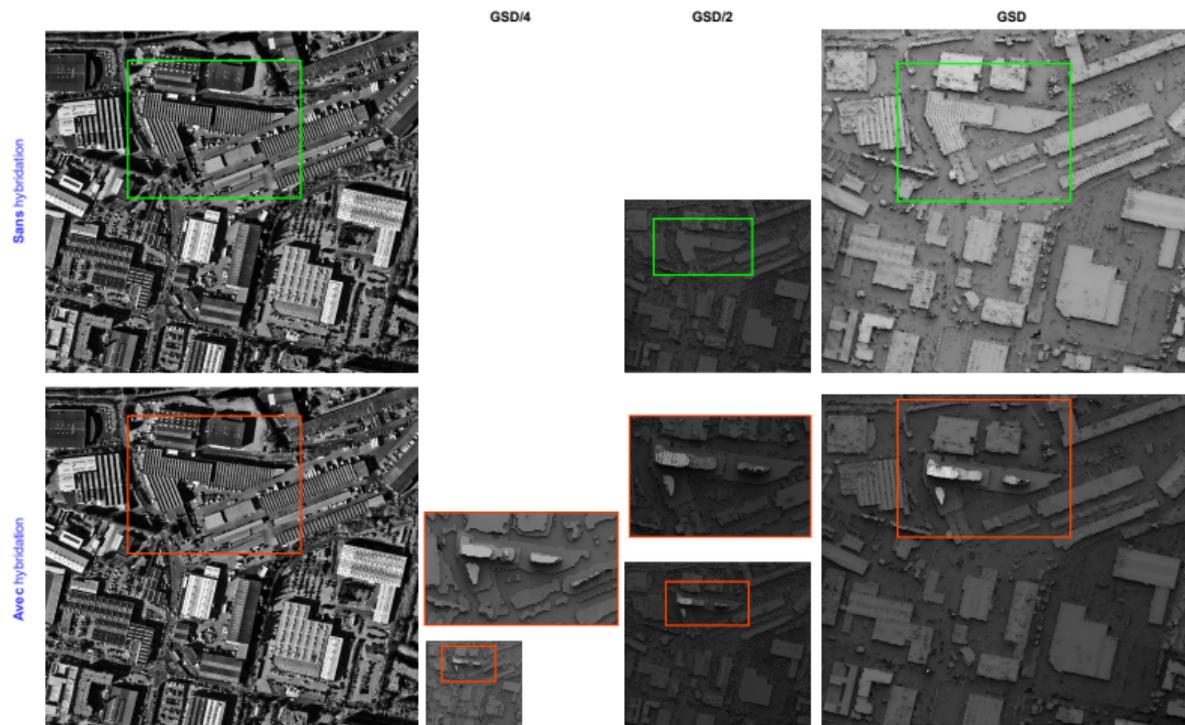
Census

NCC (3 vues)



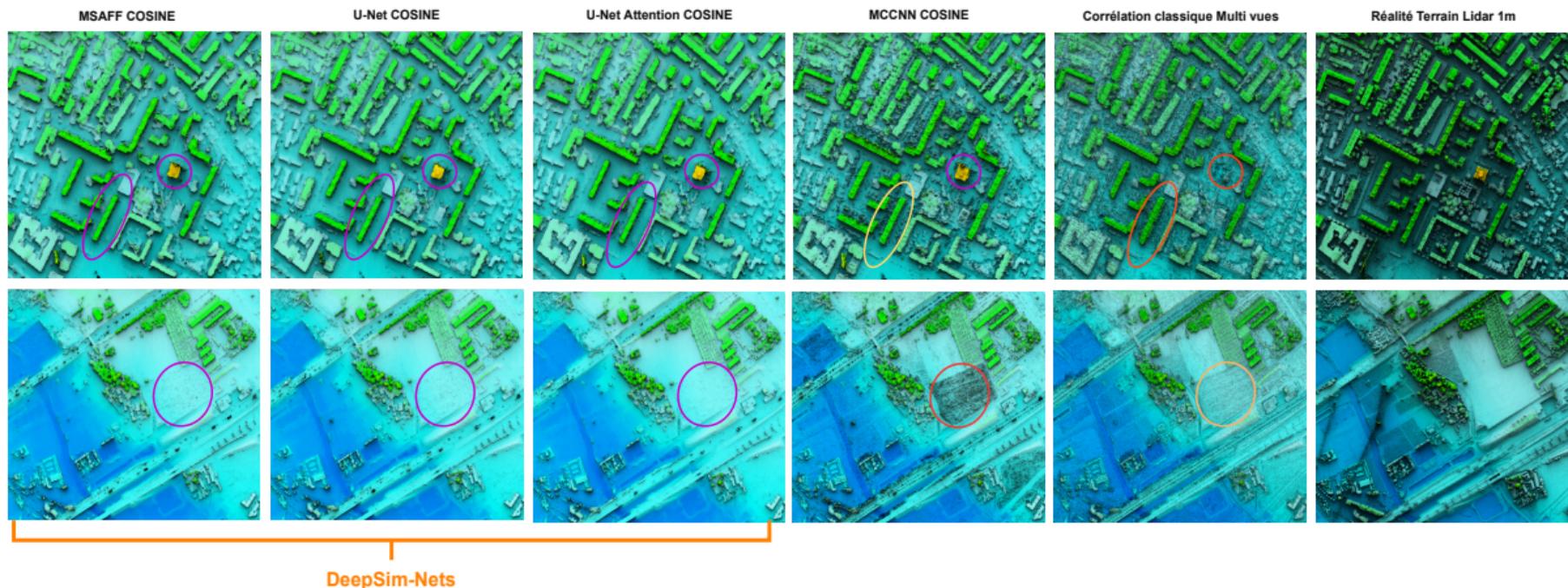
Similarité apprise vs. corrélation classique (census,ncc (3 vues)).

Approche multi-échelle: Limitations.



Propagation des erreurs d'appariement entre échelles de reconstruction 3D.

Reconstruction moins bruitée et plus fidèle % la réalité terrain.



Variantes de DeepSim-Nets comparées à MCCNN et à la corrélation (NCC).

Adaptation de domaine par fine-tuning

Adaptation aux capteurs satellitaires par finetuning.

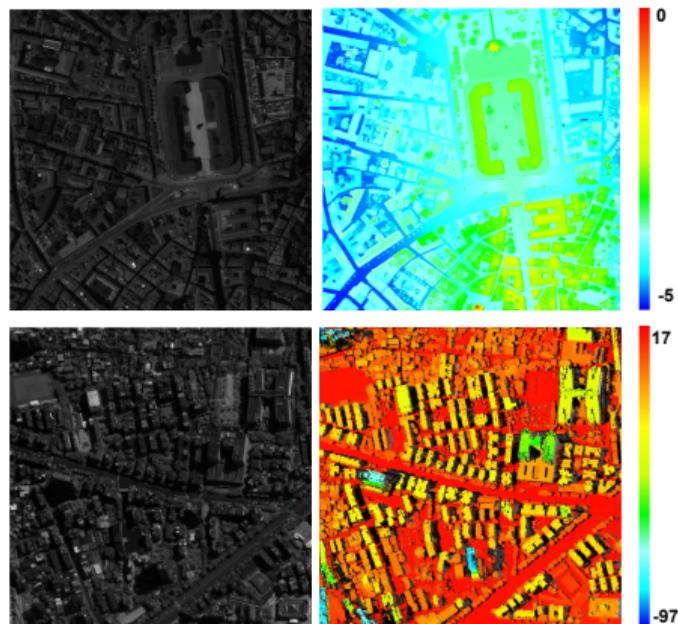


Image de référence

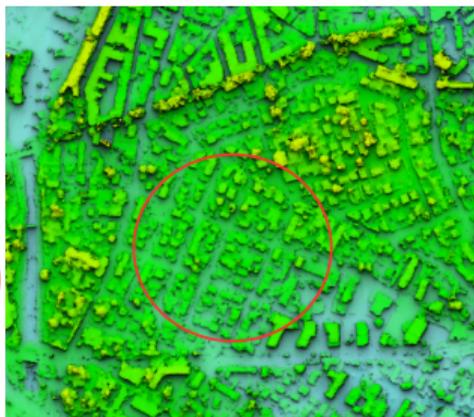
Carte de disparité

Visuel des données de finetuning.

Capteur	Nb. d'images
Pléiades (50cm)	303
GaoFen-7 (-cm)	1141

Données de finetuning

Sans finetuning



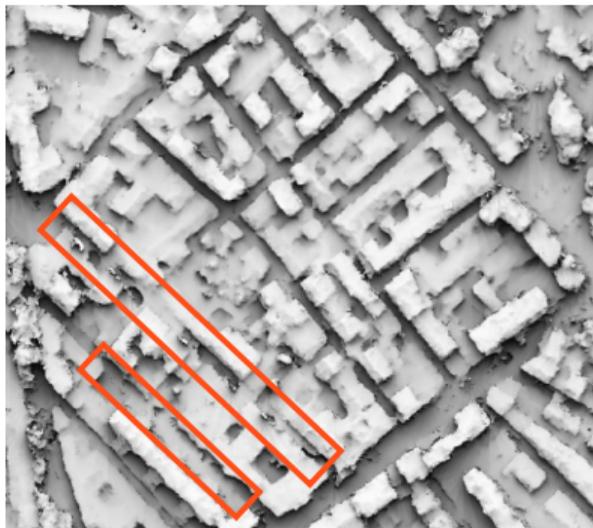
Avec finetuning



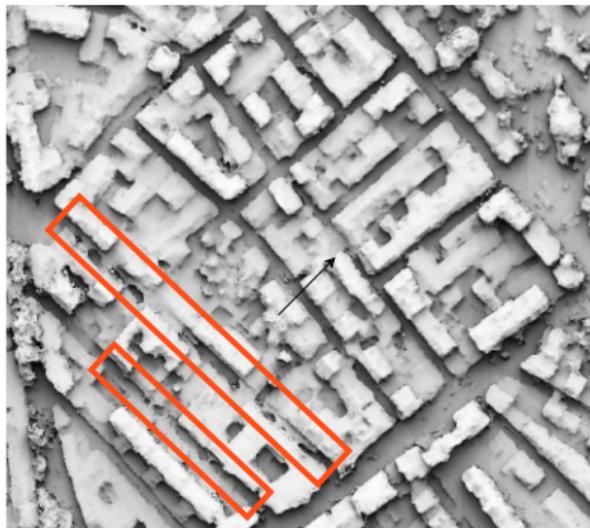
MNS de DeepSimNet(U-Net Att+MLP) avant et après finetuning.

Entraîné sur des données aériennes (6/8cm GSD)

Fusion U-Net 32 COSINE

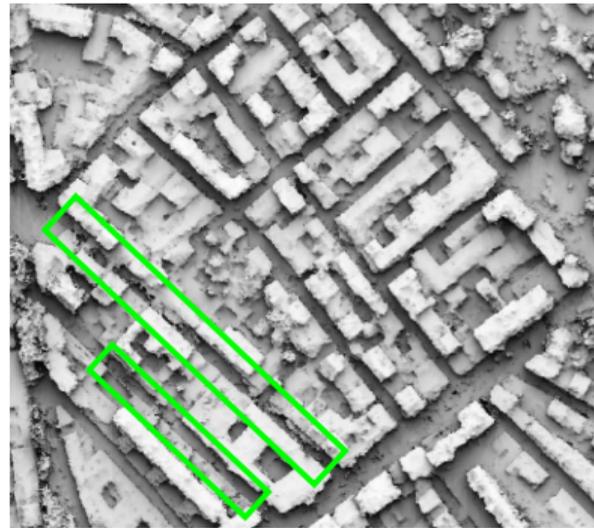


Fusion U-Net 32 COSINE+MLP



Fine-tuned sur (GaoFen, Pléiades (50cm GSD))

Fusion U-Net 32 MLP

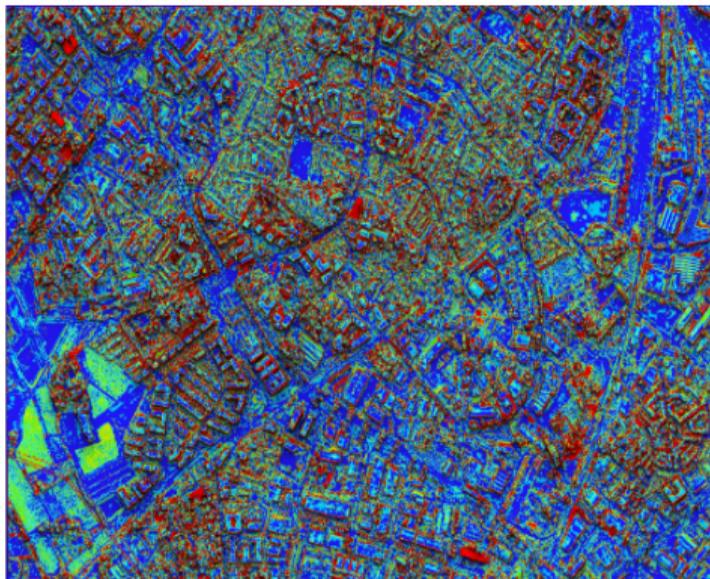


Évaluation qualitative de la séparation toiture/ruelle (quelques pixels)

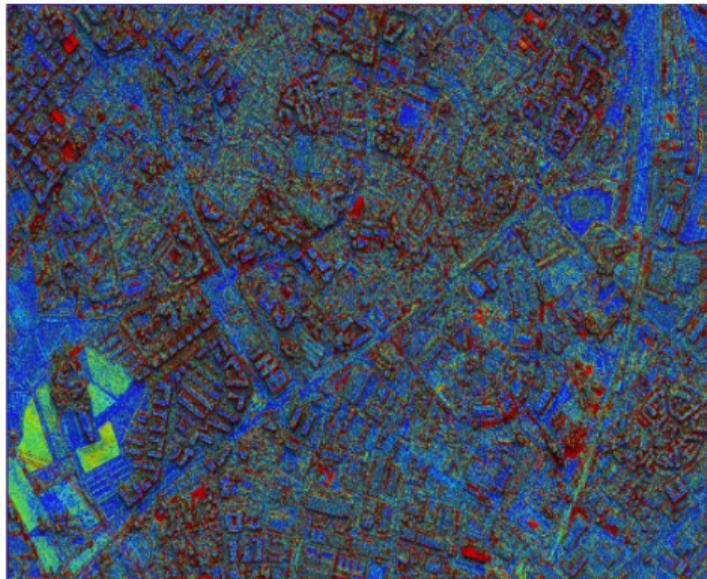
Bilan

- ▶ Approche globale de calcul de similarité dense entre tuiles épipolaires.
- ▶ Affectation de l'**ensemble** des pixels d'une image de référence.
- ▶ Approche hybride permettant de reconstruire fidèlement les surfaces homogènes par le biais de l'apprentissage ensembliste.
- ▶ Bonne généralisation sur les données PNEO.
- ▶ Amélioration des rendus par fine-tuning sur des données satellitaires Pléiades (50cm) et GaoFen-7.
- ▶ Importance de la vérité terrain (synchrone, à la résolution des images)

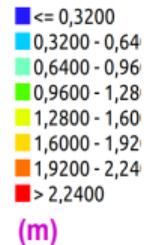
Cartes de différences absolues % à la réalité terrain



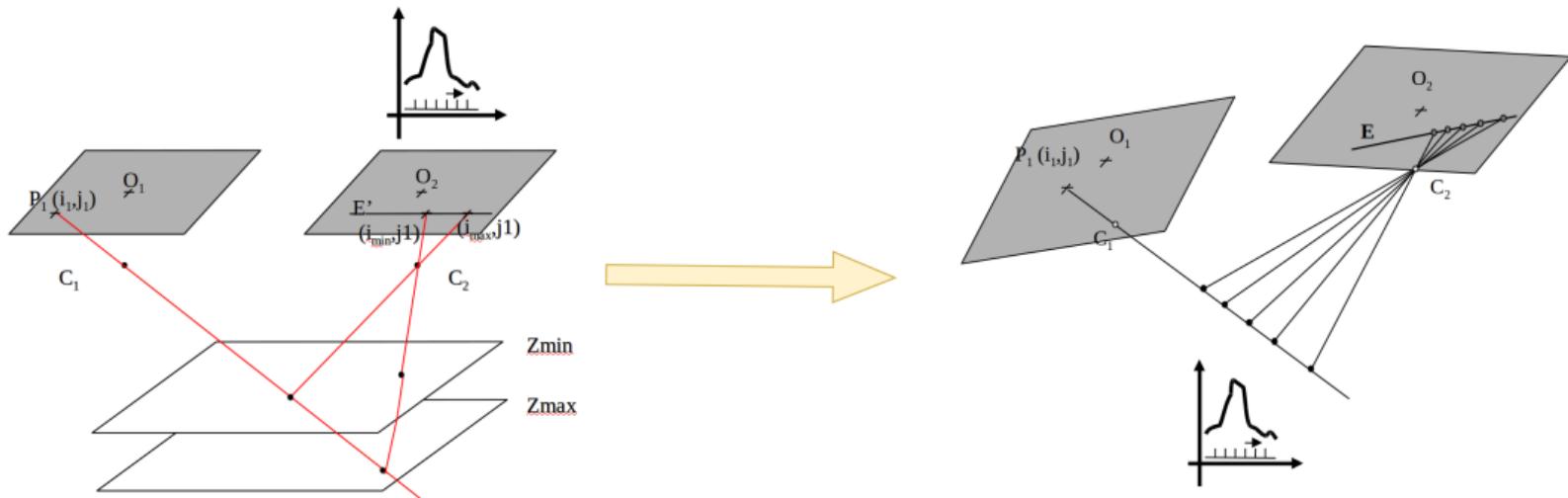
DeepSimNet



Corrélation multi vues



Perspectives



Transfert de la similarité de la géométrie épipolaire à la géométrie image.

Propriétés de la similarité

- ▶ agnostique à la géométrie sous réserve de garder le même voisinage d'un pixel.

Ce qu'on peut faire

- ▶ Appliquer des transformations géométriques connues a priori pour garantir les mêmes performances en géométrie image (≥ 3 images).

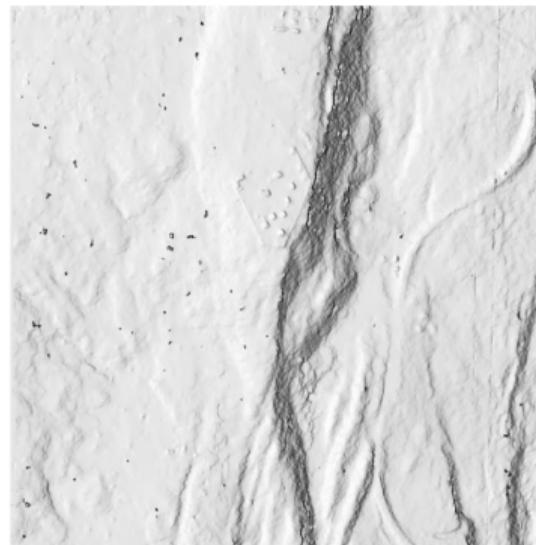
5 images Pléiades: GSD 50cm



NCC 5x5



MC-CNN [2]



DeepSim-Nets (OURS)

Modèles numériques de surface.

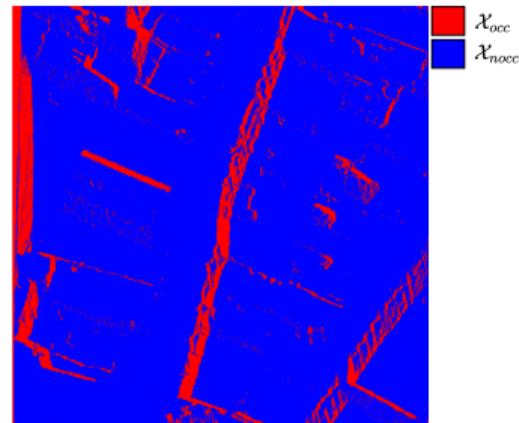
- [1] M. PIERROT DESEILLIGNY AND N. PAPRODITIS, “MULTIRESOLUTION AND OPTIMIZATION-BASED IMAGE MATCHING APPROACH : AN APPLICATION TO SURFACE RECONSTRUCTION FROM SPOT6-HRS STEREO IMAGERY,” VOL. 36 (1/W41), 1-5, 2006.
- [2] J. ZBONTAR AND Y. LECUN, “STEREO MATCHING BY TRAINING A CONVOLUTIONAL NEURAL NETWORK TO COMPARE IMAGE PATCHES,” IN *JMLR*, 2016.
- [3] H. HIRSCHMÜLLER, “STEREO PROCESSING BY SEMIGLOBAL MATCHING AND MUTUAL INFORMATION,” IN *IEEE TPAMI*, 2008.
- [4] S. ROY AND I. COX, “A MAXIMUM-FLOW FORMULATION OF THE N-CAMERA STEREO CORRESPONDENCE PROBLEM,” IN *ICCV*, 1998.
- [5] M. CHEBBI, E. RUPNIK, M. PIERROT-DESEILLIGNY, AND P. LOPES, “DEEPSIM-NETS: DEEP SIMILARITY NETWORKS FOR STEREO IMAGE MATCHING,” IN *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 2097–2105.

Autosupervisé & Contrastif

$$\begin{aligned}
 \mathcal{L}_{3All} &= \mathcal{L}_{3nocc} + \mathcal{L}_{3occ} \\
 &= \sum_{(i,j) \in \mathcal{X}_{nocc}} \mathcal{O}(\mathcal{S}_{-}^{i,j} - \mathcal{S}_{+}^{i,j} + m, 0) \\
 &\quad + \sum_{(i,j) \in \mathcal{X}_{occ}} \mathcal{O}(\mathcal{S}_{1-}^{i,j} + \mathcal{S}_{2-}^{i,j}, 0),
 \end{aligned} \tag{1}$$

Terme de pénalisation de la similarité en zones occultées.

- ▶ $\mathcal{S}_{1-}^{i,j} = \langle \mathcal{X}_{1-}^{i,j}, \mathcal{X}_{occ}^{i,j} \rangle$
- ▶ $\mathcal{S}_{2-}^{i,j} = \langle \mathcal{X}_{2-}^{i,j}, \mathcal{X}_{occ}^{i,j} \rangle$
- ▶ $\mathcal{X}_{occ}^{i,j}$: représentation de référence située dans une zone occultée (Épipolaire 1).
- ▶ $\mathcal{X}_{1-}^{i,j}$ et $\mathcal{X}_{2-}^{i,j}$ des représentations "samplées" dans la carte des représentations de l'épipolaire 2.



Définition des régions non-occultées/occultées.