Analyse Grande Echelle de Données 3D

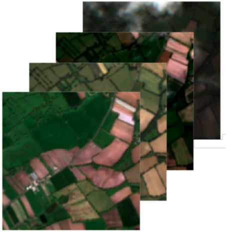
LabeX Bezout
Math/info pour la ville
2022

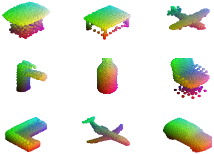
Loic Landrieu

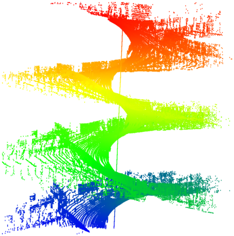
IGN - LASTIG - ENSG - UGE

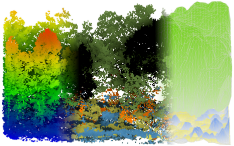
Présentation

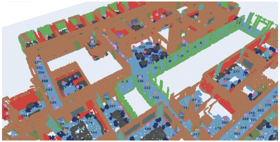
- **Loic Landrieu**: chercheur en apprentissage / vision à l'IGN.
- **3D data**: indoor, urbain, inventaire forestier
- **3D Dynamique**:conduite autonome (ANR ReADy3D)
- **Séries Temporelles**: occupation des sols
- **Optimization:** problèmes inverses, graphes
- Program Chair d'ISPRS Congress

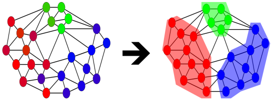










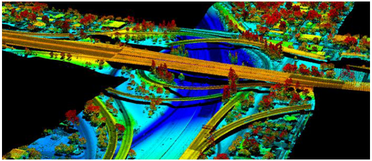


Plan

Presentation Layout

Nuage de Points 3D

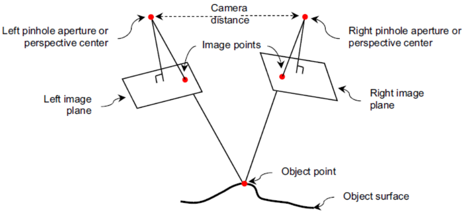
- Nuage de points 3D: $[x_i, y_i, z_i] \in \mathbf{R}^{N\times 3}$
  $+$ radiométrie



Thevara etal 2019, mathworks, renishaw.fr, tuck mapping

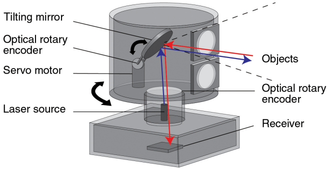
Nuage de Points 3D

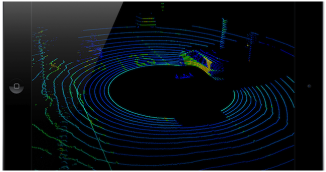
- Nuage de points 3D: $[x_i, y_i, z_i] \in \mathbf{R}^{N\times 3}$ + radiométrie
- Comment obtenir un nuage de points?:
- **Photogrammetry**: stereovision, SLAM, Surface from Motion



Thevara etal 2019, mathworks, renishaw.fr, tuck mapping

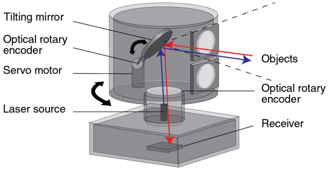
- Nuage de points 3D: $[x_i, y_i, z_i] \in \mathbf{R}^{N \times 3}$ + radiométrie
- Comment obtenir un nuage de points?:
- **Photogrammetry**: stereovision, SLAM, Surface from Motion
- **LiDAR**: active sensor with lasers

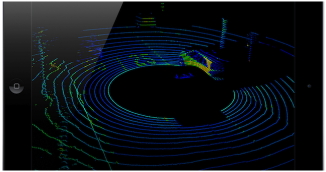




Thevara etal 2019, mathworks, renishaw.fr, tuck mapping

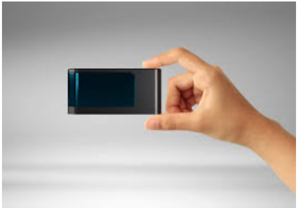
- Nuage de points 3D: $[x_i, y_i, z_i] \in \mathbf{R}^{N\times 3}$ + radiométrie
- Comment obtenir un nuage de points?:
  - **Photogrammetry**: stereovision, SLAM, Surface from Motion
  - **LiDAR**: active sensor with lasers
- Avantages et inconvénients:
  - capture la géométrie, gestion des occlusions
  - capteurs spécialisés et/ou post processing, format complexe





Thevara etal 2019, mathworks, renishaw.fr, tuck mapping

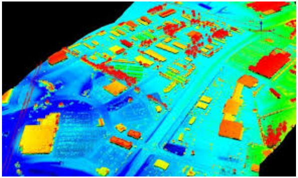
- **Accessibilité** des données 3D:
- méthodes stéréo performante
- LiDAR précis & abordables



credit: velodynelidar, green car congress

Analyse 3D: Domaine Très Dynamique

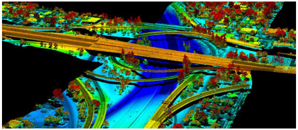
- **Accessibilité** des données 3D:
  - méthodes stéréo performante
  - LiDAR précis & abordables
- Forte **demande** d'analyse automatique.
  - Conduite autonome
  - Jumeaux numérique
  - Réalité virtuelle/augmentée
  - Surveillance environementales
  - Patrimoine digital / télé-archéologie



credit: velodynelidar, spar3d

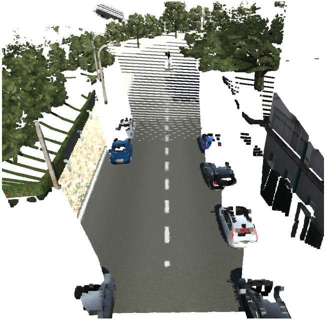
Analyse 3D: Domaine Très Dynamique

- **Accessibilité** des données 3D:
- méthodes stéréo performante
- LiDAR précis & abordables
- Forte **demande** d'analyse automatique.
- Conduite autonome
- Jumeaux numérique
- Réalité virtuelle/augmentée
- Surveillance environementales
- Patrimoine digital / télé-archéologie
- Révolution **méthodologique**
- Précision, vitese
- Passage à l'échelle reste un defis



credit: tuck mapping solutions, clearpath robotics

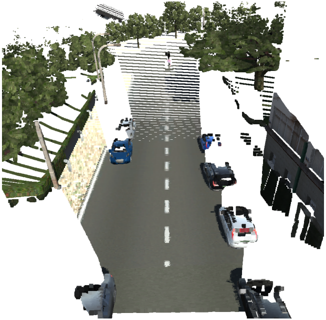
- Volume de donnée



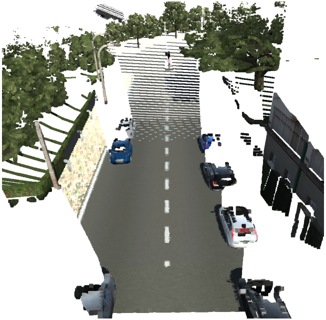
credit: Gaidon2016, Engelmann2017, Hackel2017

Complexité de l'Analyse 3D

- Volume de donnée
- Absence de grille régulière



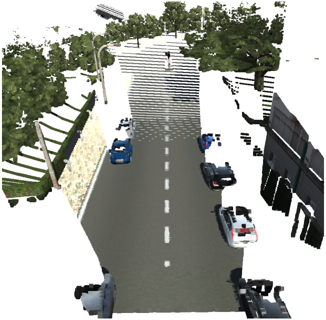
credit: Gaidon2016, Engelmann2017, Hackel2017



- Volume de donnée
- Absence de grille régulière
- Invariance par permutation

credit: Gaidon2016, Engelmann2017, Hackel2017

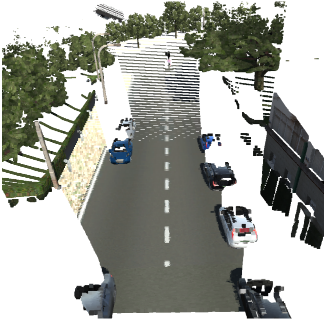
Complexité de l'Analyse 3D



- Volume de donnée
- Absence de grille régulière
- Invariance par permutation
- Parcimonie

credit: Gaidon2016, Engelmann2017, Hackel2017

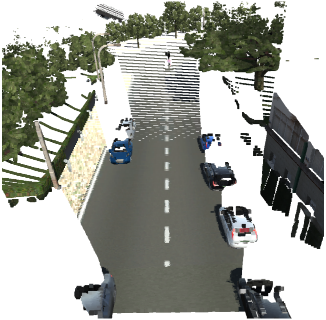
Complexité de l'Analyse 3D



- Volume de donnée
- Absence de grille régulière
- Invariance par permutation
- Parcimonie
- Densité variable

credit: Gaidon2016, Engelmann2017, Hackel2017

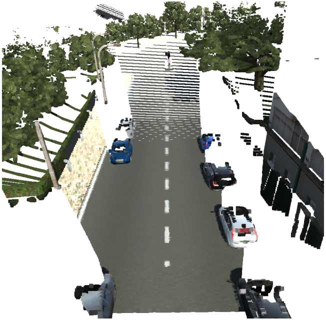
Complexité de l'Analyse 3D



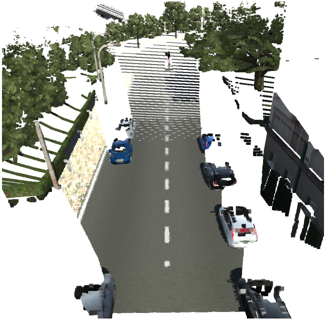
- Volume de donnée
- Absence de grille régulière
- Invariance par permutation
- Parcimonie
- Densité variable
- Artefacts d'acquisition

credit: Gaidon2016, Engelmann2017, Hackel2017

- Volume de donnée
- Absence de grille régulière
- Invariance par permutation
- Parcimonie
- Densité variable
- Artefacts d'acquisition
- Occlusions



credit: Gaidon2016, Engelmann2017, Hackel2017

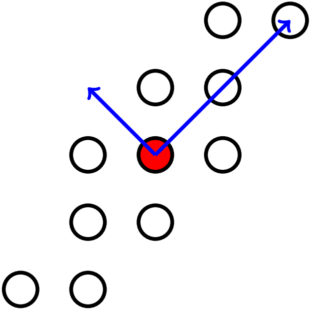


- Volume de donnée
- Absence de grille régulière
- Invariance par permutation
- Parcimonie
- Densité variable
- Artefacts d'acquisition
- Occlusions
- Apprentissage par batch

credit: Gaidon2016, Engelmann2017, Hackel2017

Presentation Layout

- **Step 1:** calcul de descripteur géométriques de points 3D:
- positions des 50 voisins
- matrice 3×3 de covariance spatiale
- valeurs propres $\lambda_1, \lambda_2, \lambda_3$
- indicateurs de dimensionalité



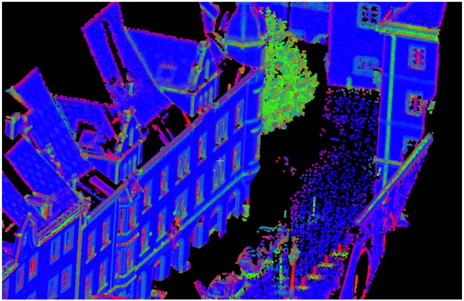
$$\text{Lin} = \frac{\sqrt{\lambda_1} - \sqrt{\lambda_2}}{\sqrt{\lambda_1}}$$

$$\text{Pla} = \frac{\sqrt{\lambda_2} - \sqrt{\lambda_3}}{\sqrt{\lambda_1}}$$

$$\text{Sca} = \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}}$$

Demantke2011,Weimann2015

- **Step 1:** calcul de descripteur géométriques de points 3D:
  - positions des 50 voisins
  - matrice 3×3 de covariance spatiale
  - valeurs propres $\lambda_1, \lambda_2, \lambda_3$
  - indicateurs de dimensionalité



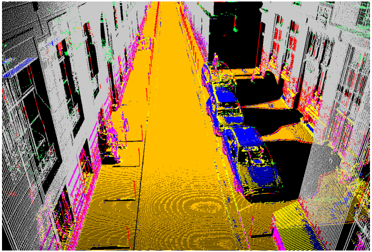
Linéarité
Planarité
Diffusion

Demantke2011,Weimann2015

credit: landrieu et. al. , JSPRS 2017

Classification de Points

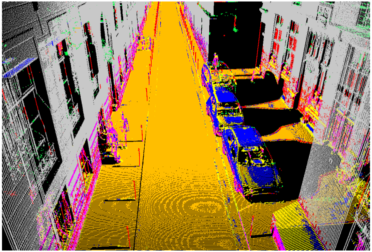
- **Step 1:** calcul de descripteur géométriques de points 3D:
- positions des 50 voisins
- matrice 3×3 de covariance spatiale
- valeurs propres $\lambda_1, \lambda_2, \lambda_3$
- indicateurs de dimensionalité
- **Step 2:**
- Concaténation avec radiométrie, elevation, etc...
- classification (RF, SVM, etc...)

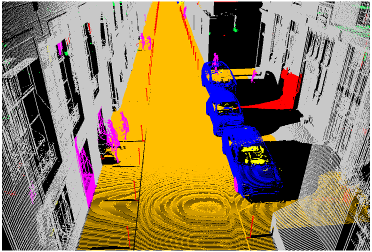


Demantke2011,Weimann2015 **credit:** landrieu et. al. , JSPRS 2017

Classification de Points

- **Step 1:** calcul de descripteur géométriques de points 3D:
  - positions des 50 voisins
  - matrice 3×3 de covariance spatiale
  - valeurs propres $\lambda_1, \lambda_2, \lambda_3$
  - indicateurs de dimensionalité
- **Step 2:**
  - Concaténation avec radiométrie, elevation, etc...
  - classification (RF, SVM, etc...)
- **Step 3:** régularisation spatiale (CRFs, MRFs, structured optimization, etc...)





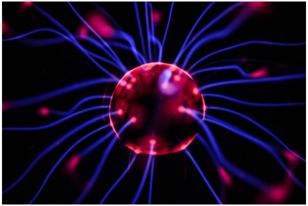
Demantke2011,Weimann2015

credit: landrieu et. al. , JSPRS 2017

Presentation Layout

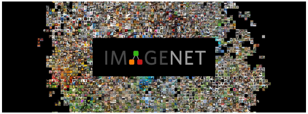
Présentation

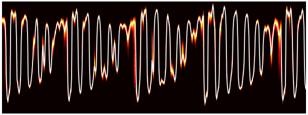
- **Deep Learning:** approche d'apprentissage machine basé sur l'usage de neurones artificiels.



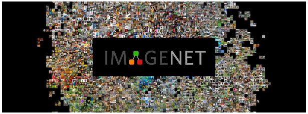
credit: towardsdatascience, deepmind, techcrunch,spar3d

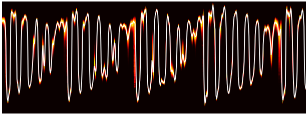
- **Deep Learning:** approche d'apprentissage machine basé sur l'usage de neurones artificiels.
- **Champs d'application:** Très efficace pour l'image, le son et le texte.

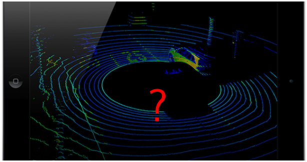




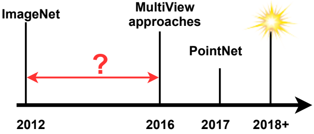
credit: towardsdatascience, deepmind, techcrunch,spar3d





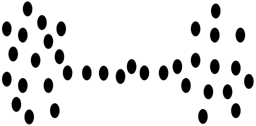


- **Deep Learning:** approche d'apprentissage machine basé sur l'usage de neurones artificiels.
- **Champs d'application:** Très efficace pour l'image, le son et le texte.
- **Et la 3D?** Est-ce vraiment plus dur?



credit: towardsdatascience, deepmind, techcrunch,spar3d

Nuages de points vus comme:



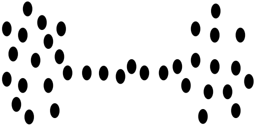
Ensemble de coordonées

PointNet, Rand-LANet

Transformers

See review by Guo *et al.* PAMI 2020

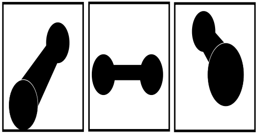
Nuages de points vus comme:



Ensemble de coordonées

PointNet, Rand-LANet

Transformers

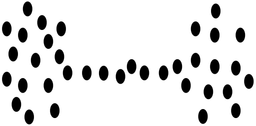


Ensemble d'image

SnapNet, MVNet

See review by Guo *et al.* PAMI 2020

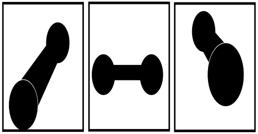
Nuages de points vus comme:



Ensemble de coordonées

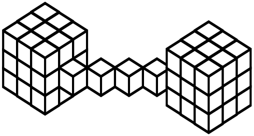
PointNet, Rand-LANet

Transformers



Ensemble d'image

SnapNet, MVNet



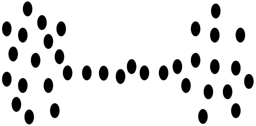
Grille de voxels

MinkowskiNet, Sparse-Conv

OctNet

See review by Guo *et al.* PAMI 2020

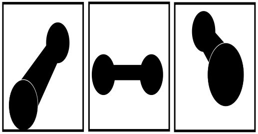
Nuages de points vus comme:



Ensemble de coordonées

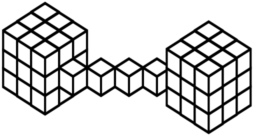
PointNet, Rand-LANet

Transformers



Ensemble d'image

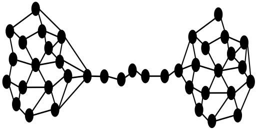
SnapNet, MVNet



Grille de voxels

MinkowskiNet, Sparse-Conv

OctNet

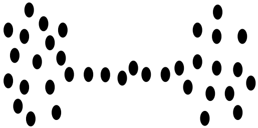


Un graphe

ECC, DGCNN

See review by Guo *et al.* PAMI 2020

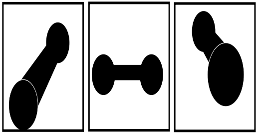
Nuages de points vus comme:



Ensemble de coordonées

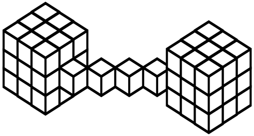
PointNet, Rand-LANet

Transformers



Ensemble d'image

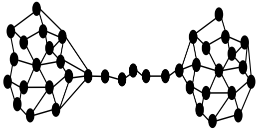
SnapNet, MVNet



Grille de voxels

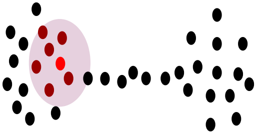
MinkowskiNet, Sparse-Conv

OctNet



Un graphe

ECC, DGCNN



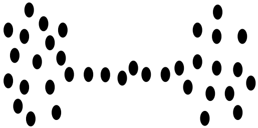
fonction dans espace continu

KPConv, ConvPoint

PointCNN

See review by Guo *et al.* PAMI 2020

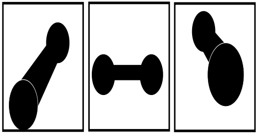
Nuages de points vus comme:



Ensemble de coordonées

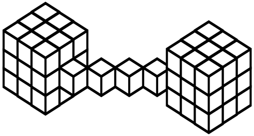
PointNet, Rand-LANet

Transformers



Ensemble d'image

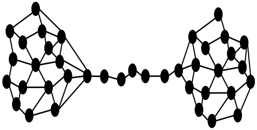
SnapNet, MVNet



Grille de voxels

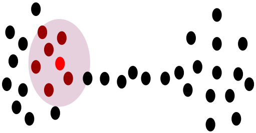
MinkowskiNet, Sparse-Conv

OctNet



Un graphe

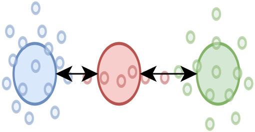
ECC, DGCNN



fonction dans espace continu

KPConv, ConvPoint

PointCNN



Graphe de superpoints

SPG, SSP

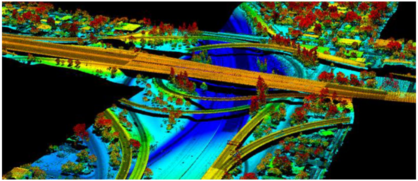
OccuSeg

See review by Guo *et al.* PAMI 2020

Presentation Layout

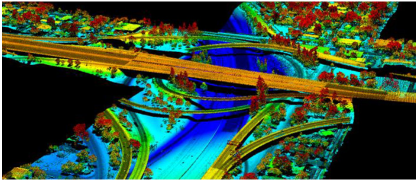
Pourquoi passer à l'échelle

- **Problème:** les méthodes précédentes sont gourmande en mémoire et compute.



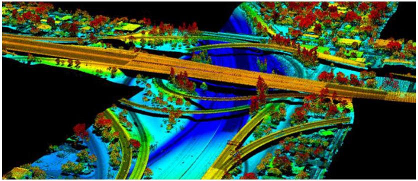
Pourquoi passer à l'échelle

- **Problème:** les méthodes précédentes sont gourmande en mémoire et compute.
- Ne peuvent être utilisées que pour quelques milliers de points.



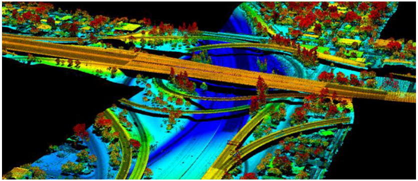
- **Problème:** les méthodes précédentes sont gourmande en mémoire et compute.
- Ne peuvent être utilisées que pour quelques milliers de points.
- **Stratégies Naïves:**

- **Sous-échantillonage**: perte d'information.



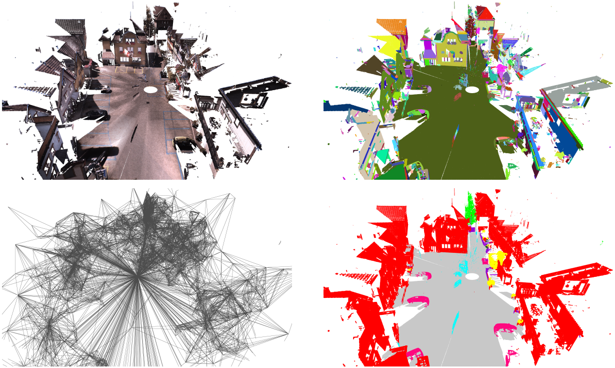
Pourquoi passer à l'échelle

- **Problème:** les méthodes précédentes sont gourmande en mémoire et compute.
- Ne peuvent être utilisées que pour quelques milliers de points.
- **Stratégies Naïves:**
- **Sous-échantillonage**: perte d'information.
- **Fenêtre glissante**: perte de la structure globale.



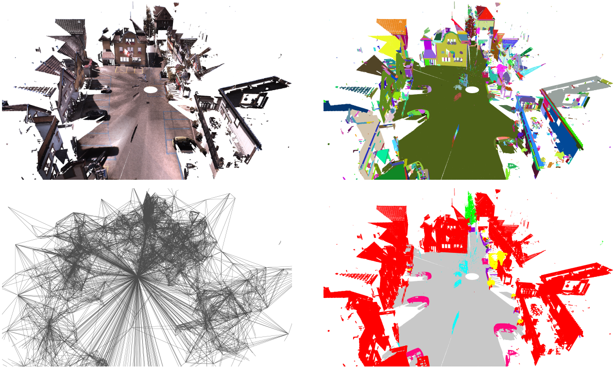
credit: tuck mapping solution

- **Observation:**
  $n_{\text{points}} \gg n_{\text{objects}}$.



Landrieu&Simonovski, CVPR 2018

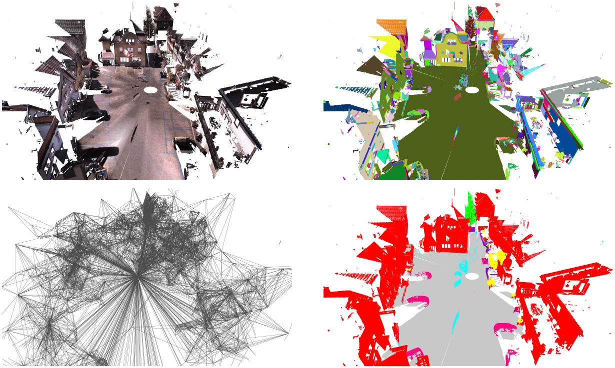
- **Observation:** $n_{\text{points}} \gg n_{\text{objects}}$.
- Partition en superpoints avec formes simples.



Landrieu&Simonovski, CVPR 2018

SuperPoint-Graph

- **Observation:** $n_{\text{points}} \gg n_{\text{objects}}$.
- Partition en superpoints avec formes simples.
- Beaucoup moins de superpoints que de points, contexte acessible par un graphe d'adjacence.

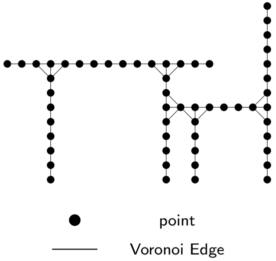


Landrieu&Simonovski, CVPR 2018

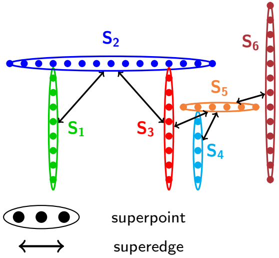
Pipeline

Step	Complexité	Algorithm
Geometric Partition en formes simples	very high 10^8 points	ℓ_0-cut pursuit
Superpoint embedding learning shape descriptors	low subsampling to 128 points	PointNet
Contextual Segmentation leveraging the global structure	very low ~ 1000 vertices	ECC with GRUs

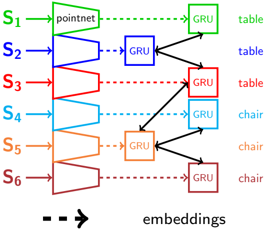
Pipeline



(a) Point cloud



(b) Superpoint graph



(c) Convolution Network

Partition

$$f^\star = \mathrm{argmin}_{f \in \mathbb{R}^{C\times m}} \sum_{i\in C} \|f_i - e_i\|^2 + \sum_{(i,j)\in E} w_{i,j}[f_i \neq f_j],$$

- $e \in \mathbb{R}^{C\times m}$: descripteurs "à la main" de la géométrie/radiométrie locales .



$$f^\star = \mathrm{argmin}_{f \in \mathbb{R}^{C \times m}} \sum_{i \in C} \|f_i - e_i\|^2 + \sum_{(i,j) \in E} w_{i,j}[f_i \neq f_j],$$

- $e \in \mathbb{R}^{C \times m}$: descripteurs "à la main" de la géométrie/radiométrie locales .
- $f^\star$: approximation constante par morceaux de e.

$$f^\star = \mathrm{argmin}_{f \in \mathbb{R}^{C\times m}} \sum_{i \in C} \|f_i - e_i\|^2 + \sum_{(i,j)\in E} w_{i,j}[f_i \neq f_j],$$

- $e \in \mathbb{R}^{C\times m}$: descripteurs "à la main" de la géométrie/radiométrie locales .
- $f^\star$: approximation constante par morceaux de e.
- Solution approximée efficacement avec ℓ_0-cut pursuit.

Landrieu & Obozinski, SIIMS 2017, Raguet & Landrieu, ICML 2018, ICML-W 2019

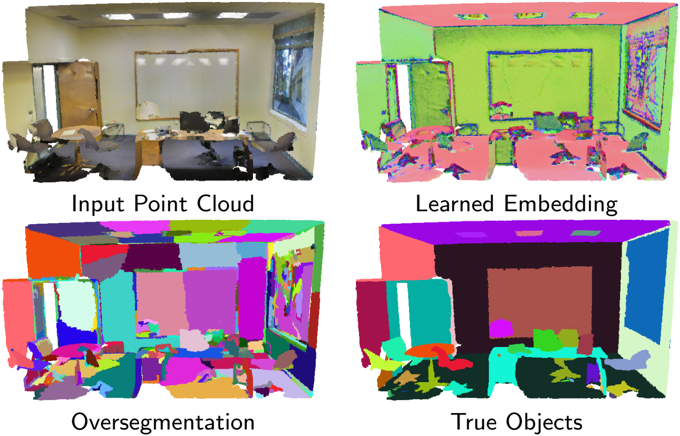
Partition

$$f^\star = \text{argmin}_{f \in \mathbb{R}^{C \times m}} \sum_{i \in C} \|f_i - e_i\|^2 + \sum_{(i,j) \in E} w_{i,j}[f_i \neq f_j],$$

- $e \in \mathbb{R}^{C \times m}$: descripteurs "à la main" de la géométrie/radiométrie locales .
- $f^\star$: approximation constante par morceaux de e.
- Solution approximée efficacement avec ℓ_0-cut pursuit.
- **Problème:** toute erreur de partition engendre une erreur de prédiction...

Landrieu & Obozinski, SIIMS 2017, Raguet & Landrieu, ICML 2018, ICML-W 2019

Apprendre à Segmenter



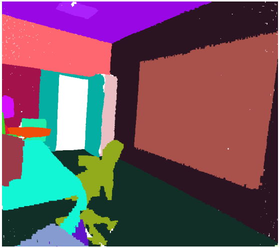
1) Descripteur de points appris pour présenter de forts contrastes à l'interface entre objets...
2) ... qui peuvent servir d'input à un algorithme de segmentation.

→ Nécessite 5 fois moins de superpoints que l'état de l'art.

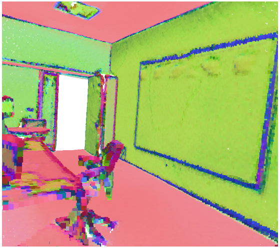
Landrieu&Boussaha, CVPR 2019

Illustration

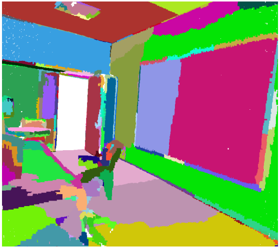
Input



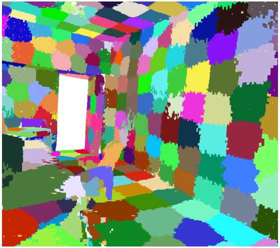
Vrais objects



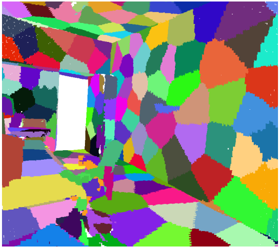
Descripteurs



Graph-LCE (ours)



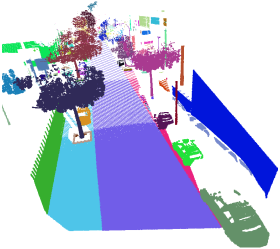
VCCS, Papon *et al.* 2013



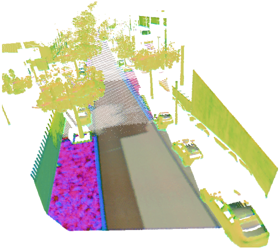
Lin *et al.* 2018

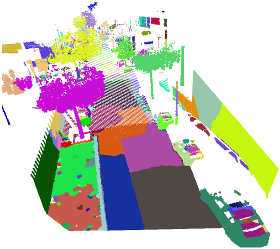
Illustration

InVrais objets

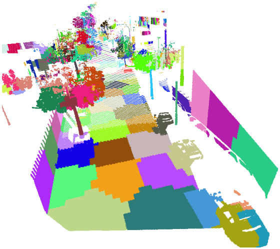


Descripteurs

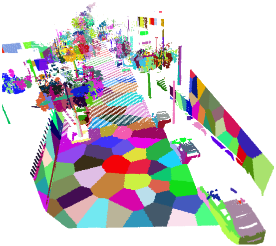




Graph-LCE (ours)



VCCS, Papon *et al.* 2013



Lin *et al.* 2018

Résultats Quantitatifs sur S3DIS:

Method	OA	mIoU
6-fold cross validation		
PointNet	78.5	47.6
Engelmann *et al.* 2017	81.1	49.7
PointNet++	81.0	54.5
<u>SPG</u>	<u>85.5</u>	<u>62.1</u>
Engelmann *et al.* 2018	84.0	58.3
PointCNN	88.1	65.4
ConvPoint 2019	88.1	68.2
<u>SSP + SPG</u>	<u>87.9</u>	<u>68.4</u>
PointSIFT	88.7	70.2
MinkowskiNet	86.0	65.9
KPConv	**88.8**	**70.6**

Table: Etat de l'art jan 2020: **OA** : Overall Accuracy, **mAcc** : average class accuracy, **mIoU**: average class Intersection over Union.

En Pratique

- **Quel algorithme utiliser?**

En Pratique

- **Quel algorithme utiliser?**

- Commencer par un algorithme classique (RF) comme baseline / nettoyage des données

En Pratique

- **Quel algorithme utiliser?**

- Commencer par un algorithme classique (RF) comme baseline / nettoyage des données
- Puis PointNet + fenêtre glissante.

En Pratique

- **Quel algorithme utiliser?**

- Commencer par un algorithme classique (RF) comme baseline / nettoyage des données
- Puis PointNet + fenêtre glissante.
- Si besoin de plus d'expressivité, MinkowskiNet/KPconv/ConvPoint.

En Pratique

- **Quel algorithme utiliser?**

- Commencer par un algorithme classique (RF) comme baseline / nettoyage des données
- Puis PointNet + fenêtre glissante.
- Si besoin de plus d'expressivité, MinkowskiNet/KPconv/ConvPoint.
- Si besoin de vitesse + structure globale: SPG.

Complementarity of 2D and 3D Modalities

- **Images:** texture and context
- **Point Clouds:** geometry
- **Objective:** combine both modalities

Robert, Vallet, Landrieu, CVPR 2022

Complementarity of 2D and 3D Modalities

- **Images:** texture and context
- **Point Clouds:** geometry
- **Objective:** combine both modalities
- Point cloud colorization: lose important information

Robert, Vallet, Landrieu, CVPR 2022

- **Images:** texture and context
- **Point Clouds:** geometry
- **Objective:** combine both modalities
- Point cloud colorization: lose important information

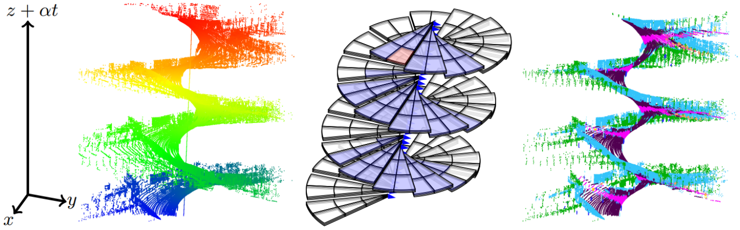
Robert, Vallet, Landrieu, CVPR 2022

Complementarity of 2D and 3D Modalities

- **Images:** texture and context
- **Point Clouds:** geometry
- **Objective:** combine both modalities
- Point cloud colorization: lose important information
- Deep Multi-View Aggregation:
  - leverage *viewing conditions*
  - from raw p.c. and images w poses
  - no mesh, depth, colorization needed
  - SOTA on S3DIS and KITTI360

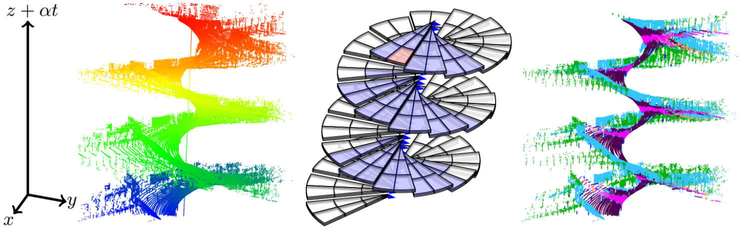
Robert, Vallet, Landrieu, CVPR 2022

Spatio-Temporal Helix



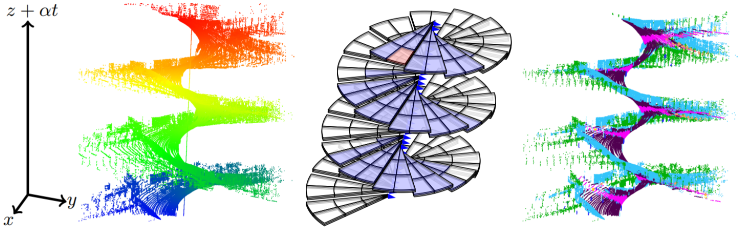
- Rotating LiDAR acquisitions have a complex spatio-temporal structure

Spatio-Temporal Helix



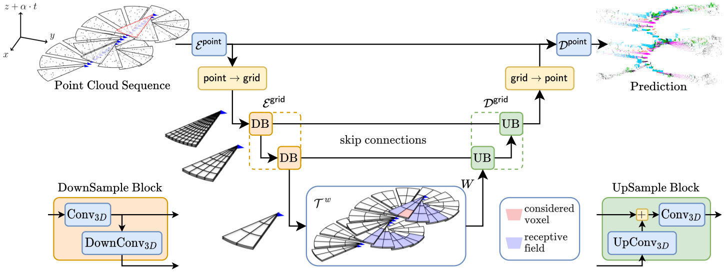
- Rotating LiDAR acquisitions have a complex spatio-temporal structure
- Most methods processes full rotations at once $\Rightarrow$ 100-200ms latency

Spatio-Temporal Helix



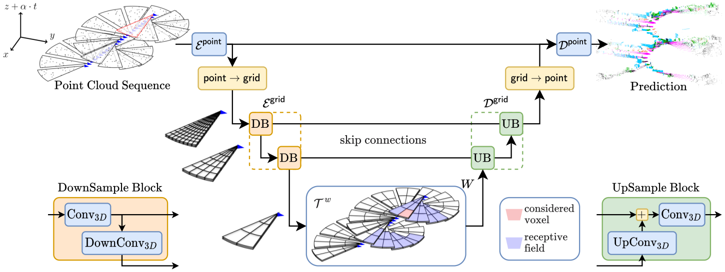
- Rotating LiDAR acquisitions have a complex spatio-temporal structure
- Most methods processes full rotations at once $\Rightarrow$ 100-200ms latency
- We split the acquisition into slices of rotation

Spatio-Temporal Helix



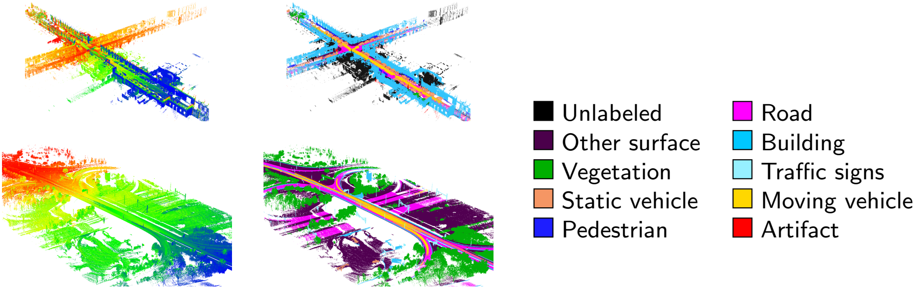
- Rotating LiDAR acquisitions have a complex spatio-temporal structure
- Most methods processes full rotations at once ⇒ 100-200ms latency
- We split the acquisition into slices of rotation
- Hybrid Spatio-Temporal Transformer:
  - Latency ÷ 5, model size ÷ 50, same accuracy

Spatio-Temporal Helix



- Rotating LiDAR acquisitions have a complex spatio-temporal structure
- Most methods processes full rotations at once ⇒ 100-200ms latency
- We split the acquisition into slices of rotation
- Hybrid Spatio-Temporal Transformer:
- Latency ÷ 5, model size ÷ 50, same accuracy
- Real-time, portable, online semantic segmentation!

Spatio-Temporal Helix



- Rotating LiDAR acquisitions have a complex spatio-temporal structure
- Most methods processes full rotations at once $\Rightarrow$ 100-200ms latency
- We split the acquisition into slices of rotation
- Hybrid Spatio-Temporal Transformer:
  - Latency $\div$ 5, model size $\div$ 50, same accuracy
  - Real-time, portable, online semantic segmentation!
- A new large-scale benchmark with 10B points and useful metadata