



HAL
open science

Recalage 3D de nuages de points par apprentissage profond

Karim Slimani, Brahim Tamadazte, Catherine Achard

► **To cite this version:**

Karim Slimani, Brahim Tamadazte, Catherine Achard. Recalage 3D de nuages de points par apprentissage profond. ORASIS 2023, May 2023, Carqueiranne, France. hal-04184149

HAL Id: hal-04184149

<https://hal.science/hal-04184149>

Submitted on 21 Aug 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.

Recalage 3D de nuages de points par apprentissage profond

Karim Slimani¹ Brahim Tamadazte¹ Catherine Achard¹

² Sorbonne Université, CNRS UMR 7222, INSERM U1150, ISIR, F-75005, Paris, France.

karim.slimani@isir.upmc.fr

Résumé

Cet article présente une nouvelle méthode utilisant l'apprentissage profond pour le recalage de nuages de points 3D. La méthode comprend un descripteur fondé sur des graphes convolutifs, un mécanisme d'attention et un algorithme de transport optimal pour estimer l'appariement des points. RANSAC est utilisé pour calculer la transformation rigide entre les deux nuages de points en fonction des correspondances prédites par le réseau. L'architecture proposée surpasse les méthodes de l'état de l'art sur la base de données ModelNet40.

Mots-Clés

Nuages de points, calcul de pose, recalage 3D, apprentissage profond, mécanisme d'attention

Abstract

This article presents a new method based on deep learning for 3D point clouds registration. The method includes a descriptor based on graph convolutions, an attention mechanism, and an optimal transport algorithm to estimate point correspondences. RANSAC is used to calculate the rigid transformation between the two point clouds based on the correspondences predicted by the network. The proposed architecture outperforms state-of-the-art methods on the ModelNet40 dataset.

Keywords

Point clouds, pose estimation, 3D registration, deep learning, attention mechanism

1 Introduction

Le recalage de points est une tâche cruciale dans l'estimation de pose 3D pour de nombreux domaines de la robotique et de la vision par ordinateur tels que la conduite autonome, l'imagerie médicale et la reconstruction d'objets. Les méthodes classiques de recalage de nuages de points impliquent deux tâches : l'appariement des points entre les ensembles de points et l'estimation des paramètres de la transformation rigide. Les algorithmes traditionnels tels que l'ICP (Iterative Closest Point) alternent entre une itération de mise en correspondance et une itération d'alignement des deux nuages de points. Les méthodes fondées sur

des réseaux de neurones telles que DCP [1] ont émergé pour répondre à cette problématique. Elles consistent à encoder chaque point et son entourage à l'aide d'un descripteur appris, puis à utiliser ces vecteurs pour propager une information locale et globale entre les deux ensembles de points. L'appariement est effectué à partir des similarités dans l'espace des descripteurs. Pour estimer les paramètres de transformation rigide, les approches end-to-end intègrent une décomposition singulière différentiable au réseau [1].

2 Méthodologie

Cette section détaille les grandes lignes du réseau proposé, qui comprennent la mise en place de notre descripteur, l'estimation de l'appariement et le calcul de la transformation nécessaire au recalage des nuages de points.

2.1 Descripteur

Afin de rendre les points plus discriminants et invariants face aux transformations rigides, une combinaison de deux encodages est proposée dans ce papier pour construire le descripteur : Dynamic Graph CNN [2] et un encodage des normales à la surface afin de capturer le voisinage local de chaque point.

Dynamic Graph CNN. Le but de ce premier descripteur est d'intégrer explicitement les caractéristiques géométriques autour de chaque point en considérant celui-ci comme le noeud d'un graphe qui le relie à ses k plus proches voisins. Un perceptron multi-couches vient ensuite construire un vecteur descripteur pour chaque noeud en appliquant des convolutions sur les coordonnées spatiales de ce même noeud et tous les autres noeuds du graphe. Afin de rendre l'opération plus robuste aux variations qui peuvent survenir dans le voisinage local, un max-pool parmi les k sorties de chaque couche est appliqué. Les graphes sont mis à jour avant chaque perceptron multi-couches en cherchant les voisins dans l'espace du descripteur prédit par le perceptron le précédent. Dans chaque couche c , le descripteur \mathbf{f}_i^c est mis à jour par la fonction suivante :

$$\mathbf{f}_i^c = \max_k \left([h_\omega^c(\mathbf{f}_i^{c-1}, \mathbf{f}_k^{c-1}); \forall k \in \mathcal{K}_i] \right) \quad (1)$$

où \mathcal{K}_i est l'ensemble des k noeuds voisins, h_ω^c sont les fonctions non-linéaires paramétrées par les poids ω au niveau de la couche c du réseau.

Encodage fréquentiel des normales à la surface. Estimées par une analyse en composantes principales, les variations des normales entre les points peuvent être exploitées afin de rendre le descripteur plus discriminant. On se propose d'encoder ces variations dans des vecteurs constitués de fonctions sinusoïdales de différentes fréquences. Ce choix est motivé par les propriétés de continuité et de différentiabilité de ces fonctions qui permettent au réseau de mieux extrapoler l'encodage pour des valeurs différentes de celles vues en phase d'apprentissage.

2.2 Mécanisme d'attention

Une fois le descripteur appliqué à chaque point des deux ensembles source et d'arrivée, un mécanisme d'attention permet de propager le contexte local et global entre les points d'un même ensemble et entre les deux ensembles. Pour ce faire, deux types d'attention sont utilisés : auto-attention et attention croisée. Dans chaque couche d'auto-attention, un graphe est construit autour de chaque point où il est relié à tous les autres points du même nuage. Tandis que dans les couches d'attention croisée, le graphe relie chaque point aux points du deuxième ensemble. Le descripteur de chaque point est mis à jour par la formule :

$$\mathbf{f}_i = \sum_{j=1}^M \alpha_{i,j} (\mathbf{f}_j \mathbf{W}^v) \quad (2)$$

avec $\alpha_{i,j}$ sont les poids d'attention donnés par :

$$\alpha_{i,j} = \text{softmax}_j \left(\frac{(\mathbf{f}_i \mathbf{W}^q)(\mathbf{f}_j \mathbf{W}^k + \mathbf{n}_{i,j} \mathbf{W}^n)^T}{\sqrt{D_{desc}}} \right) \quad (3)$$

où \mathbf{f}_i et \mathbf{f}_j sont les vecteurs descripteurs des points \mathbf{x}_i et \mathbf{x}_j , $\mathbf{n}_{i,j}$ est l'encodage du gradient des normales entre ces mêmes points, \mathbf{W}^q , \mathbf{W}^k , \mathbf{W}^n et \mathbf{W}^v représentent des matrices de projection dont les paramètres sont appris par le réseau et D_{desc} est la dimension du descripteur.

2.3 Appariement et estimation de pose

Les descripteurs prédits par le mécanisme d'attention permettent d'estimer la probabilité de correspondance entre les points. Pour ce faire, les probabilités associées aux points \mathbf{x}_i et \mathbf{x}_j sont données par le produit scalaire de leurs descripteurs et définissent une matrice de confiance \mathbf{C} :

$$C_{i,j} = \langle \mathbf{f}_i, \mathbf{f}_j \rangle \quad (4)$$

Une normalisation selon les colonnes et les lignes est effectuée par l'algorithme itératif de Sinkhorn pour obtenir une matrice d'appariement $\bar{\mathbf{C}}$ où la somme de toutes les valeurs selon chaque colonne et chaque ligne est égale à un. Les appariements sont établis en considérant que deux points \mathbf{x}_i et \mathbf{x}_j sont correspondants si et seulement si la probabilité $\bar{C}_{i,j}$ est la valeur maximale le long de la ligne i

et de la colonne j de la matrice $\bar{\mathbf{C}}$. Parmi toutes les correspondances estimées, les couples de points ayant les k^c plus grandes probabilités sont utilisés pour estimer la transformation par l'algorithme itératif RANSAC.

3 Validation

Afin d'évaluer les performances de la méthode proposée, le tableau 1 montre les résultats d'une comparaison aux techniques de l'état de l'art sur la base de données ModelNet40.

TABLE 1 – Évaluation de l'erreur en rotation (R) mesurée en degrés et en translation (t) mesurée en mètres

Method	Données Propres		Données Bruitées	
	RMSE(R)	RMSE(t)	RMSE(R)	RMSE(t)
ICP	33.683	0.293	33.067	0.294
DCP-V2	06.709	0.027	06.883	0.028
PRNET	03.199	0.016	04.323	0.017
R-PtHop	01.660	0.014	-	-
VRNet	<u>00.982</u>	<u>0.006</u>	03.615	0.010
WsDesc	<u>01.187</u>	<u>0.008</u>	<u>03.500</u>	<u>0.006</u>
Ours	00.412	0.002	01.810	0.004



FIGURE 1 – Illustration de quelques exemples de recalage effectués sur des données propres par la méthode proposée

4 Conclusion

La méthode introduite dans ce papier a montré des résultats prometteurs sur ModelNet40, à la fois sur des données propres et sur des données bruitées. De plus, l'apprentissage de la mise en correspondance permet de réduire le temps de calcul de RANSAC en utilisant 500 itérations maximum.

Références

- [1] Y. Wang and J. M. Solomon, "Deep closest point : Learning representations for point cloud registration," in *IEEE Int. Conf. on Comput. Vision*, 2019.
- [2] Y. Wang, Y. Sun, Z. Liu, *et al.*, "Dynamic graph cnn for learning on point clouds," *ACM Trans. on Graphics*, 2019.