

## SLAM LiDAR pour cartographie mobile

Jean-Emmanuel DESCHAUD

Un SLAM LiDAR est un algorithme de localisation et de cartographie en simultané basé principalement sur un capteur de type LiDAR, c'est-à-dire des données 3D géométriques, à l'inverse d'un SLAM caméra basé sur des données photométriques. L'objectif du SLAM est à la fois de créer une carte de l'environnement et de localiser le capteur dans cette même carte (capteur qui peut être porté à la main par une personne, ou positionné sur un véhicule, un robot...). Cet article expose de façon synthétique le fonctionnement d'un SLAM LiDAR avec une architecture générale et détaille plus précisément l'algorithme ICP (pour Iterative Closest Point) dont tous les SLAM les plus récents sont des variantes. Nous verrons aussi que le SLAM est encore un des rares domaines où l'IA (ou plutôt l'apprentissage machine) n'atteint pas encore les précisions des méthodes géométriques de type ICP. Enfin, nous verrons, à travers un défi annuel lancé par le groupe HILTI, que les méthodes SLAM LiDAR sont maintenant capables d'obtenir des précisions de quelques centimètres sur une trajectoire, même dans des conditions difficiles d'acquisition.

### MOTS-CLÉS

SLAM, LiDAR, cartographie, mobile, ICP

1,3 million de distances par seconde avec la même précision. Ces LiDARs restaient encombrants et lourds avec un poids d'environ 12 kg pour le Velodyne HDL64. De nos jours, nous pouvons obtenir un LiDAR comme celui de la marque Hesai, le modèle XT32, qui génère 640 000 points par seconde avec une précision de 2 cm pour un poids de seulement 1 kg.

En parallèle, avec l'évolution des capteurs, les algorithmes de SLAM (*Simultaneous Localization And Mapping*) ont aussi beaucoup progressé, le SLAM étant un domaine de recherche très actif. Les premiers SLAM LiDAR étaient basés sur un monde en 2D considéré comme plat avec une estimation de pose<sup>1</sup> du capteur de seulement trois degrés de liberté (position x, position y et angle de direction  $\theta$ ). Une application très répandue de ce type de SLAM 2D est la

Les capteurs LiDARs pour systèmes mobiles ont énormément évolué ces vingt dernières années. Les investissements pour le véhicule autonome ont permis la miniaturisation des capteurs et la diminution de leur coût. En 2005, un LiDAR 2D de marque

SICK de modèle LMS221 ne permettait d'obtenir que 14 000 points par seconde avec une précision d'environ 4-5 cm. La marque Velodyne, avec son premier modèle LiDAR 3D Velodyne HDL64, a bouleversé le marché en 2010 avec un capteur capable de mesurer

<sup>1</sup> Une pose est un terme utilisé en robotique qui signifie "élément de localisation et d'orientation".

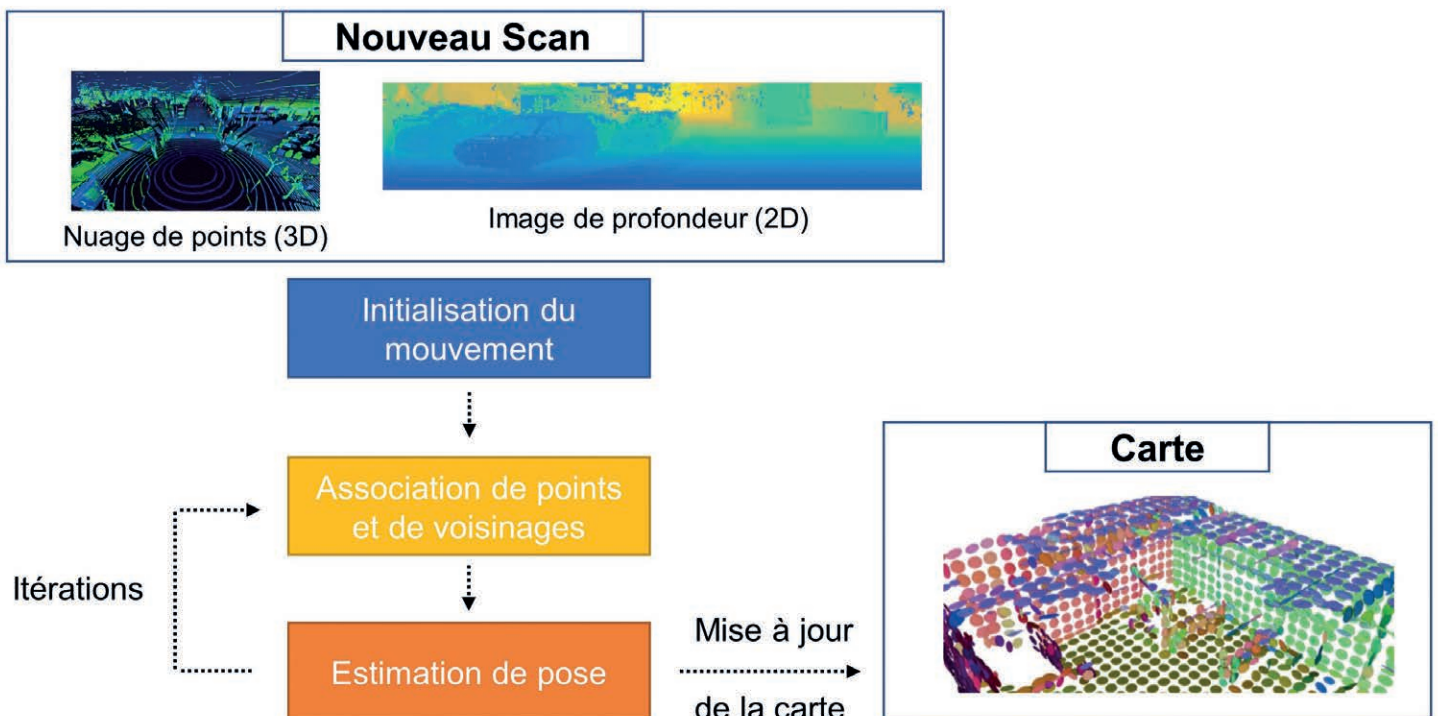


Figure 1. Architecture générale d'un SLAM LiDAR.

localisation utilisée par les robots aspirateurs. Actuellement, les algorithmes de SLAM considèrent un monde en 3D, sans hypothèse de sol plat, avec une estimation de pose du capteur de six degrés de liberté.

Ces évolutions permettent d'élargir les applications potentielles du SLAM LiDAR, notamment de la cartographie mobile dans tout type d'environnement (intérieur, extérieur, urbain, forestier...). Avant d'évoquer les méthodes les plus récentes, nous allons décrire comment se décompose un SLAM LiDAR, puis nous verrons l'évolution des algorithmes depuis la première méthode de recalage appelée ICP apparue en 1992 et qui est, encore aujourd'hui, le cœur de la plupart des algorithmes de SLAM LiDAR.

Comme on peut le voir sur la *figure 1*, un SLAM basé sur un capteur LiDAR va se décomposer en trois étapes importantes : initialisation du mouvement, association de voisinages et estimation de poses. Avant la première étape, les données, appelées scan, peuvent être reçues sous deux formes : un nuage de points (liste de points avec positions  $x, y, z$ ) ou sous forme d'une image de profondeur (une image avec une profondeur  $d$  pour chaque pixel). Le scan sous forme 3D permet d'avoir des méthodes plus précises, mais plus lentes, alors qu'un scan sous forme d'image de profondeur permet d'avoir des méthodes plus rapides, mais souvent moins précises. À partir des données d'un nouveau scan (qui représente une rotation complète à 360 degrés des tirs laser pour les LiDARs mécaniques comme le Velodyne HDL64 ou le Hesai XT32), il faut initialiser le mouvement, c'est-à-dire avoir une première estimation de la pose de ce nouveau scan. Cette initialisation va dépendre de l'application et surtout de l'utilisation du capteur (porté à la main, placé sur un robot quatre roues, sur un robot bipède...). On utilise généralement un modèle à vitesse constante qui fonctionne dans de nombreuses situations, et qui considère que le déplacement entre le scan  $n$  et le scan  $n-1$  est identique au déplacement entre le scan  $n-1$  et le scan  $n-2$ . Les deuxième et troisième étapes

d'un SLAM (association de voisinages et estimation de poses) sont liées et se font par itérations (nous développerons ci-dessous ces deux étapes importantes avec l'algorithme ICP). Une fois que la pose du nouveau scan  $n$  est estimée, on peut mettre à jour la carte qui peut se présenter sous différentes formes (nuage de points, voxels, surfels...). Une pratique répandue est de mettre la carte sous forme de nuage de points formé par l'accumulation des scans précédents (en se limitant à un nombre fixe de scans précédents et en sous-échantillonnant le nuage pour éviter d'avoir trop de points).

Nous n'allons pas le détailler dans cet article, mais il existe, en plus de cette architecture générale du SLAM, des méthodes (par optimisation d'un graphe de poses) pour améliorer la trajectoire lorsque l'on revient à un endroit déjà vu (appelé fermeture de boucle).

Le cœur du SLAM 3D est ainsi constitué des étapes d'association de voisinages et d'estimation de poses. La plupart des méthodes actuelles sont des variantes d'une méthode appelée ICP (*Iterative Closest Point*) qui date de 1992 (Besl and McKay, 1992). L'association de points pour ICP consiste à prendre pour chaque point ( $p_i$ ) du scan  $n$ , le point le plus proche ( $q_i$ ) de la carte (en considérant la carte sous forme de nuage de points). L'estimation de la pose (paramétrisée par une matrice de rotation  $R$  et une translation  $t$ ) se fait par minimisation d'une erreur de distance entre les correspondances de points :

$$\min_{R,t} \sum_i \|Rp_i + t - q_i\|^2$$

La résolution de ce problème de minimisation se fait directement par décomposition en valeurs singulières (Besl and McKay, 1992). La nouvelle pose estimée ( $R$  et  $t$ ) est appliquée au scan  $n$ , ce qui va changer l'association de voisinages (les  $q_i$ ) et ainsi amener à un nouveau problème de minimisation. Cette procédure est appliquée itérativement jusqu'à convergence du scan (c'est-à-dire jusqu'à arriver à un scan qui ne bouge presque plus d'une itération à une autre). ICP est une méthode géométrique qui fonctionne bien quand

on dispose d'une bonne initialisation, ce qui est le cas dans l'application SLAM avec la temporalité des scans (acquisition en continu de données par le LiDAR), d'où l'importance de l'étape d'initialisation dans la *figure 1*.

Les méthodes SLAM LiDAR les plus récentes se basent pour la plupart encore sur ICP et nous présentons deux méthodes publiées dont les codes sont disponibles en open source : KISS-ICP (Vizzo et al., 2023) et CT-ICP (Dellenbach et al., 2022).

KISS-ICP (*Keep It Small and Simple ICP*) de (Vizzo et al., 2023) est une méthode de l'Université de Bonn avec un code disponible sur Github : <https://github.com/PRBonn/kiss-icp>. Le principe est de rester proche de l'algorithme original ICP pour ne pas multiplier le nombre de paramètres. Cela permet pour un même jeu de paramètres à KISS-ICP de fonctionner avec un LiDAR sur un véhicule en environnement routier, sur un Segway dans un campus, avec un LiDAR solid-state porté à la main ou porté par un drone...

CT-ICP (*Continuous-Time ICP*) de (Dellenbach et al., 2022) est une méthode développée au Centre de Robotique de Mines Paris – PSL, avec un code aussi disponible en open source sous Github : [https://github.com/jedeschaud/ct\\_icp](https://github.com/jedeschaud/ct_icp). La contribution principale de CT-ICP est d'être plus précis que la méthode de base ICP qui n'estime qu'une seule pose par scan. Or, cela néglige les mouvements qui ont lieu pendant l'acquisition d'un scan (rotation de 360 degrés des tirs laser). Pour cela, CT-ICP cherche à estimer deux poses par scan, une pose de début de scan et une pose de fin de scan. Cela définit ainsi un problème d'optimisation avec 12 degrés de liberté. Pour que le problème de minimisation converge, on rajoute des contraintes de continuité intra-scan et de discontinuité inter-scan (permettant de tenir compte des mouvements brusques dans la trajectoire). CT-ICP nécessite d'ajuster plus de paramètres que KISS-ICP, mais permet d'obtenir une trajectoire et de reconstruire une carte avec plus de précision. On peut voir sur la *figure 2* un exemple de carte construite par CT-ICP en utilisant seulement un capteur Velodyne

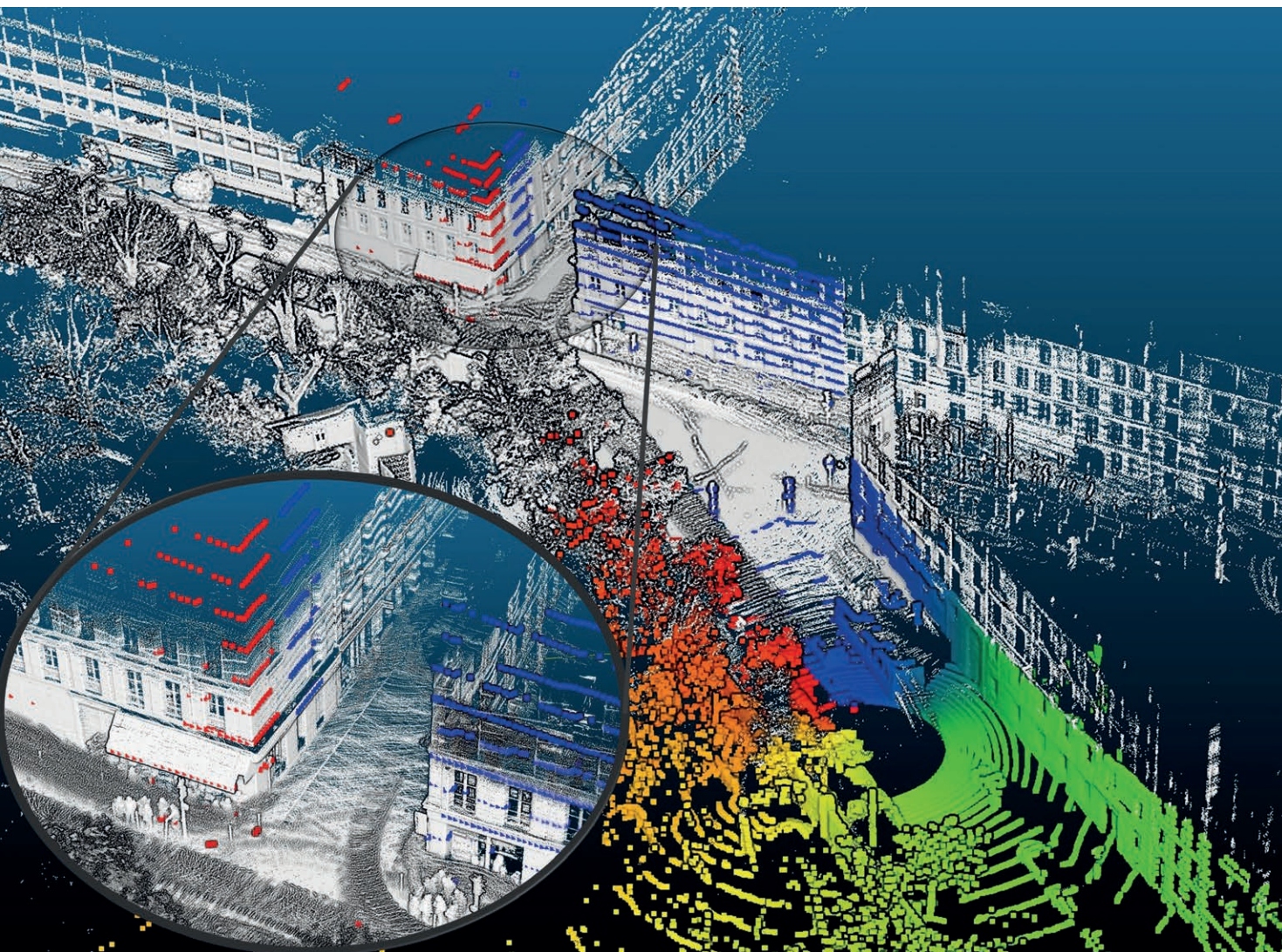


figure 2. Le SLAM LiDAR CT-ICP en environnement extérieur urbain (en couleur, le scan courant ; en blanc la carte courante sous forme de nuages de points). CT-ICP permet d'exploiter la continuité de l'acquisition pour affiner le recalage, visible en bas à gauche.

HDL32 fixé sur le toit d'un véhicule dans une acquisition autour du Jardin du Luxembourg dans le centre de Paris. On peut observer la qualité de la donnée produite alors que le système n'utilise aucun autre capteur.

Avec l'explosion de l'IA, le SLAM LiDAR basé apprentissage machine, et notamment apprentissage profond, a fait son apparition récemment avec des méthodes comme DeepLO (Cho et al., 2020). L'idée de DeepLO (*Deep LiDAR Odometry*) est d'utiliser un encodeur de type réseau convolutionnel sur les deux images de profondeur des scans  $n$  et  $n-1$ , puis, à partir des descripteurs calculés sur ces images, d'utiliser un réseau de type PoseNet qui permet de prédire une pose relative entre les deux images de profondeur. Ce type de méthode arrive à apprendre de façon non supervisée, c'est-à-dire sans nécessité d'avoir une vérité

terrain de trajectoire sur les séquences d'apprentissage. Elle repose sur une fonction de perte<sup>2</sup> géométrique de type ICP. Finalement, ces méthodes sont pour l'instant très efficaces quand la séquence de test est très proche de la séquence d'apprentissage (par exemple, dans le même environnement avec le même capteur), mais elles généralisent très mal.

Nous avons montré cela dans (Dellenbach et al., 2021) : une méthode aussi simple qu'ICP avec une bonne initialisation permet de dépasser les méthodes comme DeepLO sur des séquences non vues pendant l'apprentissage. PoseNet est un réseau permettant de prédire des poses quand

<sup>2</sup> Une fonction de perte est utilisée dans le domaine de l'intelligence artificielle et va quantifier l'écart entre les prévisions du modèle et la vérité terrain pour permettre d'ajuster les paramètres du modèle.

les données sont très proches de celles vues pendant l'entraînement. À partir de données acquises par un capteur LiDAR fixé sur le toit d'un véhicule, nous avons montré qu'en utilisant des séquences pour l'entraînement où le véhicule roule entre 0 et 60 km/h, le réseau PoseNet est incapable de prédire des poses relatives plus grandes lors d'une séquence de test, par exemple sur autoroute en roulant à 110 km/h. Cette limite qui n'a pas encore été résolue restreint fortement les applications potentielles. Ainsi, les SLAM géométriques basés sur des variantes d'ICP restent les méthodes les plus performantes actuellement.

Il est difficile de quantifier précisément la qualité d'un SLAM. Le groupe HILTI (entreprise dans le bâtiment), intéressé par les possibilités du SLAM mobile dans des sites de construction, lance chaque année depuis 2021 un

défi SLAM. Les jeux de données qui ont été créés lors de ces défis sont tous disponibles en ligne à l'adresse suivante : <https://hilti-challenge.com/>. Nous discuterons ici du défi de l'édition 2022 qui a été organisé lors d'un workshop de la conférence ICRA (*International Conference on Robotics and Automation* [IEEE]) en mai 2022 à Philadelphie. Les résultats de cette édition, ainsi que le protocole d'acquisition des données, sont décrits dans la publication (Zhang et al., 2023).

Ce défi vise toutes les solutions de SLAM (logiciels industriels ou algorithmes académiques) basées sur un ou plusieurs capteurs disponibles sur un système portable (avec capteurs caméra, LiDAR, IMU). Le système est porté à la main dans des conditions réelles difficiles (faible luminosité, masquages temporaires des capteurs vision, brusques rotations, environnements avec effet tunnel...). Toutes les équipes industrielles ou académiques ont pu soumettre les résultats de leur trajectoire calculés avec leur solution SLAM sur un site en ligne. Pour qualifier les différentes trajectoires soumises par les participants, les organisateurs ont fait un relevé topographique (avec un scanner Z+F Imager 5016) de tous les environnements. Des mires ont ensuite été placées sur des points marqués au sol avec un relevé précis de leur position 3D. Enfin, lors de l'acquisition mobile, le porteur du système posait au sol le système sur les points marqués d'une croix, points qui ont servi de points de contrôle pour mesurer la qualité des trajectoires. Les relevés ont donné, après recalage, un positionnement des scans à 95 % avec une incertitude inférieure à 3 mm.

Un exemple de séquence du défi HILTI SLAM 2022 est la séquence *Sheldonian Theatre* qui est un scan mobile en intérieur du *Sheldonian Theatre* dans l'Université d'Oxford. La séquence fait 6 min avec un porteur qui va se déplacer dans le grand théâtre, mais aussi dans les combles, les petits escaliers étroits... Il s'agit d'une séquence considérée comme la plus difficile de l'édition 2022 et la plus représentative des défis que l'on peut avoir en cartographie mobile en intérieur.

Lors de cette édition, il y a eu 42 soumissions de groupes académiques ou industriels. Tout d'abord, les organisateurs ont été obligés de revoir le système de notation, car ils ne s'attendaient pas à d'aussi bons résultats. En effet, le meilleur SLAM obtient une précision moyenne par point de 2,07 cm sur l'ensemble des séquences. Une autre conclusion de ce défi est que les SLAM basés LiDARs sont beaucoup plus précis que les SLAM basés caméra (2,07 cm de précision pour le meilleur SLAM LiDAR contre 25,36 cm de précision pour le meilleur SLAM caméra). Enfin, sur la séquence la plus difficile de ce défi, le *Sheldonian Theatre*, la meilleure méthode SLAM a permis d'obtenir une précision moyenne d'environ 4 cm sur l'ensemble du parcours. Ce niveau de précision montre que les méthodes SLAM LiDARs mobiles sont matures pour être exploitées en conditions réelles, permettant d'augmenter la productivité d'acquisition en intérieur comme en extérieur. ●

## Contact

Jean-Emmanuel DESCHAUD  
[jean-emmanuel.deschaud@minesparis.psl.eu](mailto:jean-emmanuel.deschaud@minesparis.psl.eu)  
 Enseignant-chercheur au Centre de Robotique de Mines Paris - PSL  
 Spécialisé en Vision 3D pour systèmes autonomes

## Bibliographie

(Besl and McKay, 1992) Besl P. J. and McKay, N. D. (1992), "A method for registration of 3-D shapes," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239-256.  
 (Cho et al., 2020) Cho, Y., Kim G. and Kim, A. (2020), "Unsupervised Geometry-Aware Deep LiDAR Odometry," in IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020, pp. 2145-2152.  
 (Dellenbach et al., 2021) Dellenbach, P., Deschaud, J.-E., Jacquet, B. and Goulette, F. (2021), "What's in My LiDAR Odometry Toolbox?" in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 2021, pp. 4429-4436.

(Dellenbach et al., 2022) Dellenbach, D., Deschaud, J.-E., Jacquet, B. and Goulette, F. (2022), "CT-ICP: Real-time Elastic LiDAR Odometry with Loop Closure", in International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 2022, pp. 5580-5586.

(Vizzo et al., 2023) Vizzo, I., Guadagnino, T., Mersch, B., Wiesmann, L., Behley, J. and Stachniss, C. (2023), "KISS-ICP: In Defense of Point-to-Point ICP - Simple, Accurate, and Robust Registration If Done the Right Way," in IEEE Robotics and Automation Letters, vol. 8, no. 2, pp. 1029-1036.

(Zhang et al., 2023) Zhang, L., Helmlinger, M., Fu, L. F. T., Wisth, D., Camurri, M., Scaramuzza, D. and Fallon, M. (2023), "Hilti-Oxford Dataset: A Millimeter-Accurate Benchmark for Simultaneous Localization and Mapping," in IEEE Robotics and Automation Letters, vol. 8, no. 1, pp. 408-415.

## ABSTRACT

**Keywords:** SLAM, LiDAR, mapping, mobile, ICP

A LiDAR SLAM is a Simultaneous Localization And Mapping algorithm based mainly on a LiDAR-type sensor, i.e. 3D geometric data, unlike a camera SLAM based on photometric data. The objective of SLAM is both to create a map of the environment and to locate the sensor in this same map (sensor which can be carried in the hand by a person, or positioned on a vehicle, a robot, etc.). This article summarizes the pipeline of a LiDAR SLAM with a general architecture and details more precisely the ICP algorithm (for Iterative Closest Point) of which all the most recent SLAMs are variants. We will also see that SLAM is still one of the few areas where AI (or rather machine learning) does not yet reach the precision of ICP-type geometric methods. Finally, we will see through an annual challenge launched by the HILTI group that LiDAR SLAM methods are now able to obtain accuracies of a few centimeters on a trajectory even under difficult acquisition conditions.