ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-3/W1-2022
14th GeoInformation for Disaster Management (Gi4DM 2022), 1–4 November 2022, Beijing, China

# A BRIEF OVERVIEW OF THE CURRENT STATE, CHALLENGING ISSUES AND FUTURE DIRECTIONS OF POINT CLOUD REGISTRATION

Nathan Brightman, Lei Fan*

Department of Civil Engineering, Xi'an Jiaotong-Liverpool University, 111 Renai Road, Suzhou, 215123, P.R. China

**Commission III**

**KEY WORDS:** Point Cloud; Registration; Deep Learning; Convolutional Neural Network; LiDAR; Laser Scanning

**ABSTRACT:**

Point cloud registration is the process of transforming multiple point clouds obtained at different locations of the same scene into a common coordinate system, forming an integrated dataset representing the scene surveyed. In addition to the typical target-based registration method, there are various registration methods that are based on using only the point cloud data captured (i.e. cloud-to-cloud methods). Until recently, cloud-to-cloud registration methods have generally adopted a coarse-to-fine optimisation process. The challenges and limitations inherent in this process have shaped the development of point cloud registration and the associated software tools over the past three decades. Based on the success of applying deep learning approaches to imagery data, numerous attempts at applying such approaches to point cloud datasets have received much attention. This study reviews and comment on recent developments in point cloud registration without using any targets and explores remaining issues, based on which recommendations on potential future studies in this topic are made.

## 1. INTRODUCTION

Point cloud data is a very basic form of data, a set of points representing a group of objects and the space between them. As such, it finds utility in a broad range of applications at vastly different scales, from the very large, such as geographic survey, through to the very small, such as microbiology or particle physics. Point cloud registration is a basic step in many point cloud processing pipelines. It is the process of aligning two or more 3D point clouds collected at different locations of the same scene. Registration enables point cloud data to be transformed into a common coordinate system, forming an integrated dataset representing the scene surveyed. There are various registration methods available, such as those reliant on targets being placed in the scene before data capture, and others based on using only the data captured [Fan et al., 2015].

The motivation behind point cloud registration may be broadly split into two categories: the desire to build models based on multiple point clouds (Cai, 2021), or the desire to know the relative position or the pose of one point cloud with respect to another (Fan, 2020). These different motivations place different emphasis on the registration process, for example, either towards achieving high precision or high speed.

Different applications and data acquisition methods influence the importance of key factors in the registration process, such as the degree of overlap between point clouds, the type of transformation (e.g. rigid or non-rigid) needed to complete the registration, the level of error and noise present in the data.

Registration between source and target point clouds is commonly a two-step process: (1) establishing 3D-3D point correspondences between the source and target, and (2) finding the optimal transformation between the source and the target. Considering rigid transformations (a combination of rotation and translation), the optimal transformation is usually considered to be one that minimises the total Euclidean distance between all point correspondences.

This study briefly reviews and discusses the recent key developments in point cloud registration without using any targets. Until recently, point cloud registration methods have generally been centred upon the use of a coarse-to-fine optimisation strategy, the best-known element of which is Iterative Closest Point (ICP). The challenges and limitations inherent in this process have shaped the development of point cloud registration and the associated software tools over the past three decades. Based on recent success for deep learning methods applied to 2D image data, attempts at applying these approaches to 3D data sets have received much attention. The fusion of more recent deep learning methods and conventional optimisation approaches is the source of much research and progress. We review the state of the art in both approaches and highlight various remaining issues in this subject.

## 2. POINT CLOUD DATA

### 2.1 File Formats

There is a wide range of file formats, including ASCII (XYZ, OBJ, PTX and ASC), binary (FLS, PCD and LAS), or those (e.g. PLY, FBX and E57) storing data in both binary and ASCII. Which specific one is used will depend on both the source of the data and the tools used in handling them. Whilst the underlying X-Y-Z format is very much canonical, other data is often associated with both the individual points and the point cloud as a whole. The impact the specific format will have is in terms of the data available to guide any processing: additional data available for each point (e.g. colour, intensity), and/or meta data available to, for example, provide class, time or location parameters. Apart from the likely benefits of the additional data, they also present a complicating factor if, for example, point

*Correspondence: lei.fan@xjtlu.edu.cn

clouds derived from more than one data format are to be used with each other.

### 2.2 Benchmark Datasets

In any development work aimed at improving point cloud data processes, there is a need to have access to substantial amounts of point cloud data. When datasets have been specifically collected for a particular set of experiments or real-world application, sometimes these are shared with the rest of the community. For example, if any comparison is to be made between tools and techniques, access to such common datasets is very valuable. When used in this way, these datasets are often referred to as "benchmarks". A number of valuable attempts have been made to provide publicly shared evaluation benchmarks. Typically, these cover a specific use case of point clouds and have been collected using a specific sensor. Preparing data for point cloud related experiments is a substantial task. Some of the benchmark datasets acknowledge this by providing some welcome assistance, perhaps either in the form of documentation or specific software to aid with data preparation. There are a range of available datasets, some of which are more applicable than others to particular problem types, such as cloud-to-cloud registration.

Since deep learning requires large volumes of data to facilitate training of the learning models, simulated data is also of particular value. In general terms, the objective is to closely simulate the characteristics of data collected using real sensor devices. Some key attributes of simulated point clouds to be considered include: level of random errors, interaction with reflective surface (e.g. glass, water), variable atmospheric conditions (e.g. rain, fog), point density. The broad aim is to facilitate an analysis of the performance of registration methods based on using simulated point cloud datasets. Therefore, for completeness, corresponding real-world point cloud data should then be used to evaluate the applicability of the models trained on the simulated data sets.

Several sources of point cloud data that may be considered include real-world single-sensor data (e.g. 3DMatch (Zeng et al., 2017), KITTI (Geiger et al., 2012), ETHdata (Pomerleau et al., 2012)), real-word multi-sensor data (e.g. 3DCSR (Huang et al., 2021), and synthetic (often simulated) point clouds. The synthetic point clouds may come from model-based such as ModelNet40 (Wu et al., 2015), or simulated sensors such as BLAINDER (Reitmann et al., 2021), SynthCity (Griffiths and Boehm, 2019), CARLA simulator (CARLA, 2021), LGSVL Simulator (LGSVL, 2021).

## 3. REGISTRATION METHODS

### 3.1 Typical Registration Strategy

Some key factors that may be considered for developing registration methods include: point clouds from different sensor types where different noise patterns may exist, degrees of overlap, amount of misalignment, combination of both rotation and translation errors, categories of point cloud registration problems, global and/or local, indoor structured scenes, outdoor unstructured/structured, with and without moving objects, scale of problems etc.

Most point cloud registration methods employ a coarse-to-fine strategy. In this approach, a coarse registration is first applied to find an approximate rigid transformation (a combination of rotation and translation) for a pair of point clouds. Once a coarse transformation is available, a fine registration algorithm, such as ICP, Normal Distribution Transform (NDT), or one of the more efficient variants of ICP&NDT is used to refine the final transformation.

A global registration problem is one where the aim is to align point clouds without additional information on their relative 'poses'. Most global algorithms do not lend themselves to providing precise results. In contrast, local registration algorithms perform better in this respect. As might be expected, algorithms focused on local registration are usually less effective for global problems. One issue is that they make use of local optimisation techniques that may get stuck within local minima when used against specific datasets. As a result, many registration pipelines use a global algorithm to provide an initial estimate that sets up a subsequent local process. Global registration is typically achieved using geometric features. Such feature-based registration is usually a slow process as the extraction of features can be computationally expensive. Generally, local registration approaches, do not usually employ any feature extraction.

ICP is the most studied and remains the de facto technique for local registration. Generally, ICP assumes that the point clouds are roughly aligned and aims to calculate the rigid transformation that achieves the alignment. Rather than comparing features, ICP approximates potential correspondences by looking for the closest point to each point, which is often an expensive computation process. A large number of variants of ICP have been developed, typically focusing on speeding or quality of the results. There are also local registration algorithms that do not employ such nearest-point approximation. NDT, for example, treats the point clouds as a set of Gaussians, trying to align them by finding the most probable alignment.

### 3.2 Local Registration

#### 3.2.1 ICP and Variants:

It is widely accepted that the ICP algorithm is the most widely adopted method for pairwise fine registration. Requiring a good initial transformation to bring it close to registration, ICP converges to a more optimal registration by repeatedly applying a search for point-to-point correspondences followed by a transformation calculation. LiDAR point clouds are often huge and corrupted by variations in the point density, noise, outliers (unintended points), occlusions (missing points) and partial overlaps. ICP is challenged by such LiDAR point clouds due to limited one-to-one correspondences between two point clouds. A significant body of research has been applied to variants of ICP, aiming at dealing with these challenges. Representative ones include point-to-plane (Chen and Medioni, 1992), point-to-projection (Campbell and Flynn, 2001), plane-to-plane (Segal et al., 2009). Chetverikov et al. (2002) proposed a trimmed ICP (TrICP) algorithm. At each iteration step, TrICP considers the outliers, shape defects and partial overlaps, making it more tolerant of incomplete and noisy data. Yang et al. (2013) introduced the global optimal ICP method (Go-ICP) to integrate ICP with a branch-and-bound scheme, so coarse registration is not needed. However, Go-ICP is much more time consuming than ICP and sensitive to outliers. Another focus on improving ICP is the efficiency of correspondence search, for example, a GPU accelerated way to deal with K-D Tree structures (Qiu et al., 2009).

#### 3.2.2 NDT and Variants:

The NDT algorithm is one alternative fine registration method, initially developed for 2D robotics and then 3D (Takeuchi and Tsubouchi, 2006; Nüchter, 2008). NDT handles the registration process as one of matching probability density functions (PDFs). Registration problem is transformed into a nonlinear optimisation problem where the optimal transformation is based on maximizing the similarity between the PDFs. Whereas ICP and its variants require a relatively high point density (to obtain accurate correspondences), NDT is better able to tolerate lower/variable density in point clouds. Similar to ICP, NDT has been refined and continues to be incorporated in new research and developments, for example Zhou et al. (2021).

### 3.3 Classic, Non-learning-based Global Registration

In conventional global registration approaches, geometric features are extracted using manually defined rules to form the handcrafted feature descriptors. For the handcrafted feature-based, a deeper review can be found in Han et al. (2018). Fast Point Feature Histogram (FPFH) (Rusu et al., 2009) appears to be the basis of a lot of research and being used in various works claiming state-of-the-art results. It is observed that the FPFH descriptor has been utilised in a number of places and it continues to provide state-of-the-art results with the TEASER (Yang and Carlone, 2019; Yang et al., 2020). One related approach is Fast Global Registration (FGR) (Zhou et al., 2016). FGR is commonly used as a benchmark for being a state-of-the-art global registration method. Four-points congruent set (4PCS) based registration also continues to drive a number of research efforts; one which is the basis of the current state-of-the-art results is K-4PCS (Theiler et al., 2014).

Probabilistic registration methods model the distribution of the point clouds as a density function. One key method, that adopts probability density estimation, is Coherent Point Drift (CPD). CPD-based methods use Gaussian Mixture Models (GMM) to describe a point cloud and then fit the GMM to a second point cloud by maximising the likelihood of the objective function. CPD sees applications for non-rigid transformations (i.e. deformation) in, for example, medical applications (Leong-Hoï et al., 2020). More generally, CPD provides generality, accuracy, and good robustness to noise and outliers. It continues to be improved by a number of works, including Golyanik et al. (2016) and Wang et al. (2019a).

### 3.4 Deep Learning-based Global Registration

Using handcrafted features to distinguish correspondences is highly dependent on the experience of their designers (Dong et al., 2020). As a result, their generalizability and robustness may be sub-optimal for many applications. The aim of using deep learning-based methods is to learn feature representations to achieve good performance. These methods can automatically learn more robust feature representations, and have great potentials for registering scenes with repetitive and symmetrical features and limited overlaps.

Based on the taxonomy in Dong et al. (2020), deep learning-based registration methods can be divided into three categories according to their data representations: voxels, multiviews and points. The voxelization-based and the multiview-based registration methods have been the subject of a number of research efforts but, due to issues around computational inefficiency have been largely restricted to good results with small-scale indoor datasets. However, a number of the results from these efforts have been utilised in developments focussed on point-based representations.

Applying deep learning on 3D point cloud data introduces a number of challenges. Some of these challenges include the general point cloud data characteristics such as occlusions and noises/outliers. However, more specific issues with the application of deep learning on point clouds are: (1) irregularity (i.e. the points are not evenly distributed spatially across the different regions of the scene so that some regions will be dense points and others sparse), (2) unstructured (i.e. not organised in a known pattern as would be the case for image data), and (3) unordered (i.e. point cloud of a scene is the set of points, usually stored as a list in a file. Being a set, a change in the order in which the points are stored does not reflect a change in the scene represented.).

These issues make the direct application of convolutional neural networks (CNN) difficult as they assume ordered, regular structures. Early approaches attempted to overcome these issues by converting the point cloud into a structured grid format. Providing an approach to this was the key step that PointNet (Qi et al., 2017a) represented. PointNet and PointNet++ (Qi et al., 2017b) are the pioneering methods for directly processing unordered point sets (invariance to transformations), which are the foundation for much of the recent developments, used as feature extractor in a number of works. For example, PPFNet (Deng et al., 2018) is based on and extends PointNet to provide some learning of the local geometry.

PointNetLK (Aoki et al., 2019) introduces the Lucas-Kanade algorithm into 3D point cloud registration and solves the problem iteratively with PointNet. PointNetLK represents a significant milestone in the development of a category of deep learning methods that do not directly seek to identify correspondences across the input point cloud data before proceeding. PointNetLK builds upon PointNet, using its learnable structured representation for point clouds, applying it to the task of point cloud registration. To achieve this, it utilizes a classical stereo vision technique, i.e. the Lucas & Kanade (LK) algorithm (Lucas et al., 1981). This connection was motivated by Wang et al. (2018a) that demonstrated 2D object tracking performance by treating the LK algorithm as a recurrent neural network, effectively extending a successful approach from 2 to 3 dimensions. It has provided an important stepping stone in some promising developments (Huang et al., 2020; Li et al., 2020).

PCRNet (Sarode et al., 2019), in a similar way to PointNetLK, utilizes PointNet to extract global features. In contrast to PointNetLK, for the feature alignment module, a data-driven technique is used. Two global features are concatenated, before five, fully connected layers are applied before an output layer provides the registration transformation. Compared to PointNetLK, PCRNet exhibits better generalizability, but is not robust to noises.

Like PointNet, although not specifically a registration method, dynamic graph convolutional neural network (DGCNN) (Wang et al., 2018b) is used as a component in a number of related registration pipelines, including Deep Closest Point (DCP) (Wang and Solomon, 2019b) and PRNet (Wang and Solomon, 2019c). In DGCNN, a graph is constructed in the feature space and dynamically updated after each layer of the network. A multilayer perceptron (MLP) is used as the feature learning function for each edge, and channel-wise symmetric aggregation is applied onto the edge features associated with the neighbours of each point.

DCP employs a DGCNN for feature extraction and a singular value decomposition (SVD) module (Papadopoulo and Lourakis, 2000) to calculate rotation and translation. Incorporating techniques from both computer vision and natural language processing, it is broadly based on the classic ICP pipeline whilst aiming to avoid the associated issue of converging to local solutions. As a limitation, there is an assumption of a high-degree of correspondence between the point clouds.

DCP has three stages. In the first stage, data are embedded (by a point cloud embedding network) into high-dimensional space using DGCNN to extract features. It is claimed that this improves the feature effectiveness of the matching by making the features task specific. In the second stage, an attention-based module (Vaswani et al., 2017) combined with a pointer generation layer (Vinyals et al., 2015) is used to approximate combinatorial matching, which provides a dependency term between the feature sets, i.e., one set is modified in a way that is based on the structure of the other. In the third stage, a differentiable SVD layer is used to extract the final rigid transformation. It is shown that SVD provides better results than using an MLP.

PRNet uses the network architecture DCP iteratively. Use of DCP in this way was suggested by the authors of DCP as a possible extension of their work to better handle partial overlap scenarios. Point cloud registration between clouds with only partial overlap is a much more challenging case to handle. Under such cases, the end-to-end, correspondence-free methods such as PointNetLK can perform poorly. PRNet is aimed at this problem. In PRNet a search for the key points is made by comparing the norms of the learned features, and then estimating the correspondences iteratively in a coarse-to-fine manner.

RPM-Net (Yew and Lee, 2020) illustrates a common theme in deep learning approaches. It adopts and builds upon a classical method Robust Point Matching (RPM) (Gold et al., 1998). RPM aims to avoid some of the issues with ICP by a soft assignment scheme combined with 'deterministic annealing' to gradually 'harden' the assignment. RPM-Net uses this soft assignment approach combined with a Sinkhorn layer (Sinkhorn, 1964). Sinkhorn is a mechanism that finds utility in a number of recent deep learning works, including PRNet. RPM-Net is, in many ways, similar to DCP. The authors claimed that their use of Sinkhorn normalization, enabled RPM-Net to better handle outliers and partial visibility. It also uses an iterative pipeline to achieve high precision, one of the ways that the DCP authors highlighted for further study. It also makes use of a FPFH (Rusu et al., 2009) based local feature descriptor that is referenced in many works, including PPFNet (Deng et al., 2018), which in turn can be utilised in PointNetLK. RPM-Net provides a useful comparison against DCP since it uses a similar testing methodology and datasets along with using many of the same benchmark methods, including FGR and PointNetLK.

For a more in-depth review of deep learning methods, see Zhang et al. (2020) and Huang et al. (2021). So far, the deep learning methods have proved effective for the registration of indoor and relatively small-scale outdoor point clouds (Dong et al., 2020). However, limitations on the amount of data and complexity mean that scaling to large-scale outdoor point clouds is still a barrier.

### 3.5 Issues

There is a wide range of ways to store point clouds in files. In practice, there are far more file extensions than there are fundamental differences between the file types. Some formats are versions tailored to proprietary file systems, optimised in one way or another for a particular software tool. These may bring inconveniences when data are shared. In some file types, additional information (other than coordinates) is also available. While such additional information may aid in data visualisation and processing, they could also present a complicating factor if, for example, point clouds derived from more than one data format are to be used with each other.

As point cloud data are a canonical form of 3D representation, it finds applications in a very wide range of applications with differing requirements and types of sensor collecting data. Therefore, point cloud registration is not one, clearly defined task. Factors such as the level of overlap, occlusion and degree of noise, rigid or non-rigid transformation can change the nature of the problem. Compounding this variability, as we have noted, across the different disciplines, the motivation for registration is either high precision to accurately capture reality or high-speed (with sufficient precision) to allow navigation. When attempting to compare the effectiveness of different processes and results from different sets of research, it is very difficult to accommodate the wide range of differences between application type (e.g. robotics, BIM, navigation, etc), data sets and sensor type used (e.g. RGB-D, Lidar, etc).

Based on our study of recently developed methods, there is not any substantial evidence that there are significant improvements in the effectiveness of the registration being achieved (i.e. precision). There is still a substantial reliance on well-established methods (e.g. ICP and variants, NDT and variants) for precision registration. It is often the case that these classic techniques are utilised to refine the registration efforts of other methods. However, there are clear signs that efficiency (i.e. speed) has benefited from recent developments. This possibly reflects where most of the emphasis in the research is being placed, broadly directed towards SLAM (simultaneous localization and mapping) to assist with the autonomous vehicle research efforts.

It is noted that the significant efforts have been directed at applying deep learning techniques to handling point cloud data, including registration. There are clear benefits available from employing deep learning-based feature extractors to avoid the need to invest in the development of their hand-crafted counterparts. These can be utilised now to give immediate benefits in identifying features to facilitate global registration. One example of this is the use of deep learning modules like DGCNN and PPFNet feature extractors. These are shown to be more effective than the non-learning-based feature extractors such as FPFH. However, it is also clear that deep learning techniques are still some way from being able to handle large scale projects where there are possibly billions of data points spread across hundreds of scans, as might be faced in a significant construction project, for example. The current state of the art is effectively limited to dealing with small to medium scale data sets. A note of caution is needed as it remains to be demonstrated that the deep learning approaches can generalise sufficiently and do not fall into the trap of 'over fitting' to their training data; a situation where the deep learning model, actually models the training data too well and cannot be used outside of that dataset effectively.

There is a limited number of pre-trained deep learning models for various typical real-life scenes, partially due to the lack of benchmark datasets. Pre-trained models are a useful way to distribute the result of training a deep learning model. Given the length of time (many hours or days) the training process for some models takes, it is highly desirable to store the parameters derived from the training in a file(s) to enable re-loading of the model at a later date. For example, if an author of a specific deep learning model provides a relevant pre-trained model, which matches users' experiment requirements, a significant amount of time can be saved if the model is to be used as a benchmark for future works. Providing both codes and pre-trained models helps make the work both a very good platform to study and a useful benchmark.

### 3.6 Future Research

Based on our study, the likely future research opportunities are recommended in the following.

1. The application of deep learning methods to large scale datasets in a systematic and reproducible way, which will require access to large datasets that are appropriate to a specific application area.
2. Further development of tools to facilitate the generation of simulated sensor data based on realistic environments (e.g. works such as BLAINDER and SynthCity show the way forward on this).
3. Based on more synthetic dataset generators, the application of Generative Adversarial Networks (GAN) techniques might yield benefits in delivering more realistic datasets, much like some of the remarkable results being seen with synthetic 2D imagery.
4. A greater emphasis on the inclusion of meta-data and additional point data (e.g. colour, multispectral bands) in the development of registration processes. This will be particularly relevant in the context of SLAM based capture devices.
5. For applications such as BIM (Building Information Modelling), a focus on using registration techniques to monitor changes in a scene over time could be fruitful.
6. Establishment of competent pre-trained deep learning models for typical real-life scenes of a particular application would be very useful for the future efficient utility of the models.

## ACKNOWLEDGEMENTS

The author is grateful for the financial support from Xi'an Jiaotong–Liverpool University, including Research Enhancement Fund (grant number REF-21-01-003) and Key Program Special Fund (grant number KSF-E-40).

## REFERENCES

Aoki, Y., Goforth, H., Srivatsan, R.A., Lucey, S., 2019: PointNetLK: Robust & efficient point cloud registration using PointNet. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7163–7172.

Cai, Y., Fan, L., 2021: An efficient approach to automatic construction of 3D watertight geometry of buildings using point clouds. *Remote Sensing*, 13(10), 1947.

Campbell, R.J., Flynn, P.J., 2001: A survey of free-form object representation and re-cognition techniques. *Computer Vision and Image Understanding*, 81(2), 166–210.

CARLA Simulator [Internet]. Readthedocs.io. 2021 [cited 2021 May 31]. Available from: https://carla.readthedocs.io/en/0.9.11/

Chen, Y., Medioni, G.G., 1992: Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3), 145-155.

Chetverikov, D., Svirko, D., Stepanov, D., Krsek, P., 2002: The trimmed iterative closest point algorithm. *IEEE Object Recognition Supported by User Interaction for Service Robots*, 3, 545–548.

Deng, H., Birdal, T., Ilic S., 2018: PPFNet: Global context aware local features for robust 3D point matching. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. doi.org/10.1109/CVPR.2018.00028.

Dong, Z., Liang, F., Yang, B., Xu, Y., Yufu Zang, Li, J., Wang, Y., Dai, W., Fan, H., Hyyppä, J., Stilla, U., 2020: Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark, *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 327-342. doi.org/10.1016/j.isprsjprs.2020.03.013.

Fan, L., Smethurst, J.A., Atkinson, P.M., Powrie, W., 2015: Error in target-based georeferencing and registration in terrestrial laser scanning. *Computers & geosciences*, 83, 54-64. doi.org/10.1016/j.cageo.2015.06.021

Fan, L., 2020: A comparison between structure-from-motion and terrestrial laser scanning for deriving surface roughness: a case study on a sandy terrain surface. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 1225-1229. doi.org/10.5194/isprs-archives-XLII-3-W10-1225-2020.

Geiger, A.P. Lenz, Urtasun, R., 2012: Are we ready for autonomous driving? The KITTI vision benchmark suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi.org/10.1109/CVPR.2012.6248074

Gold, S., Rangarajan, A., Lu, C.P., Pappu, S., Mjolsness, E., 1998: New algorithms for 2D and 3D point matching: pose estimation and correspondence. *Pattern Recognition*, 31(8), 1019 – 1031.

Golyanik, V., Taetz, B., Reis, G., Stricker, D., 2016: Extended coherent point drift algorithm with correspondence priors and optimal subsampling. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 1-9.* doi.org/*10.1109/WACV.2016.7477719.*

Griffiths, D.J., Boehm, J., 2019: SynthCity: A large scale synthetic point cloud. *ArXiv*: abs/1907.04758.

Han, X., Jin, J., Xie, J., Wang, M., Jiang, W., 2018: A comprehensive review of 3D point cloud descriptors. *ArXiv*: abs/1802.02297.

Huang, X., Mei, G., Zhang, J., 2020: Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences. P*roceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11366– 11374.

Huang, X., Mei, G., Zhang, J., Abbas, R., 2021: A comprehensive survey on point cloud registration. *ArXiv:* abs/2103.02690.

Leong-Hoï, A., Chambrial, A., Collet, M., Radoux, J., 2020: Non-rigid registration of 3D points clouds of deformed liver models with Open3D and PyCPD. *Proceedings of Unconventional Optical Imaging II.* doi.org/10.1117/12.2555673

LGSVL. lgsvl/simulator [Internet]. GitHub. 2021 [cited 2021 May 31]. Available from: https://github.com/lgsvl/simulator

Li, X., Pontes, J.K., Lucey, S., 2020): Deterministic PointNetLK for Generalized Registration. *ArXiv*: abs/2008.09527.

Lucas, B.D., Kanade, T., et al., 1981: An Iterative Image Registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*.

Nüchter, A., 2008: 3D Robotic Mapping: the Simultaneous Localization and Mapping Problem with Six Degrees of Freedom. Springer Publishing Company.

Papadopoulo, T., Lourakis, M. I., 2000: Estimating the Jacobian of the singular value decomposition: Theory and applications. *European Conference on Computer Vision. Springer, Berlin,* 554–570.

Pomerleau, F., et al., 2012: Challenging data sets for point cloud registration algorithms. *The International Journal of Robotics Research*, 31(14), 1705-1711.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 652–660.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems,* 5099–5108.

Qiu, D., May, S., Nüchter, A., 2009: GPU-accelerated nearest neighbor search for 3D re- gistration. *International Conference on Computer Vision Systems,* Springer, Berlin, 194–203.

Reitmann, S., Neumann, L., Jung, B., 2021: BLAINDER-A Blender AI Add-On for Generation of Semantically Labeled Depth-Sensing Data. *Sensors (Basel)*, 21(6), 2144.

Rusu, R.B., Blodow, N., Beetz, M., 2009: Fast point feature histograms (FPFH) for 3D registration. *IEEE International Conference on Robotics and Automation (ICRA)*, 3212–3217.

Sarode, V., Li, X., Goforth, H., Aoki, Y., Srivatsan, R.A., Lucey, S., Choset, H., 2019: PCRNet: Point Cloud Registration Network using PointNet Encoding. *ArXiv*: 1908.07906.

Segal, A., Haehnel, D., Thrun, S., 2009: Generalized-ICP. *Robotics: Science and Systems*, 5, 168–176.

Sinkhorn, R., 1964: A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2), 876–879.

Takeuchi, E., Tsubouchi, T., 2006: A 3-D scan matching using improved 3-D normal distributions transforms for mobile robotic mapping. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3068–3073.

Theiler, P.W., Wegner, J.D., Schindler, K., 1996, Keypoint-based 4-Points Congruent Sets – Automated marker-less registration of laser scans. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96, 149-163.

Vaswani, A., Shazeer, N, Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017: Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*, Advances, 5998–6008.

Vinyals, O., Fortunato, M, Jaitly, N., 2015: Pointer networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*, (2), 2692–2700.

Wang, C., Galoogahi, H. K., Lin, C.H., Lucey, S., 2018a: Deep-LK for efficient adaptive object tracking. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 627–634.

Wang, L., Chen, J., Li, X., Fang, Y., 2019a: Non-Rigid Point Set Registration Networks. *ArXiv*, abs/1904.01428.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2018b: Dynamic graph CNN for learning on point clouds. *arXiv*: 1801.07829.

Wang, Y., Solomon, J.M., 2019b: Deep closest point: Learning representations for point cloud registration. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* 3523–3532. doi.org/10.1109/ICCV.2019.00362

Wang, Y., Solomon, J.M., 2019c: PRNet: Self-supervised learning for partial-to-partial registration. *arXiv*: 1910.12240.

Wu, Z., et al., 2015: 3D shapenets: A deep representation for volumetric shapes. *Proceedings of the IEEE conference on computer vision and pattern recognition.*

Yang, J., Li, H., Jia, Y., 2013: Go-ICP: solving 3D registration efficiently and globally optimally. *Proceedings of the IEEE Conference on Computer Vision,* 1457–1464.

Yang, H., Carlone, L., 2019: A Polynomial-time Solution for Robust Registration with Extreme Outlier Rates. *ArXiv:* abs/1903.08588.

Yang, H., Shi, J., Carlone, L., 2020: TEASER: Fast and Certifiable Point Cloud Registration. arXiv:2001.07715.

Yew, Z.J., Lee, G.H., 2020: RPM-Net: Robust point matching using learned features. *2020 IEEE/CVF Conference on*

*Computer Vision and Pattern Recognition (CVPR)*, doi.org/ 10.1109/CVPR42600.2020.01184.

Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T., 2017: 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, DOI: 10.1109/CVPR.2017.29.

Zhang, Z. Dai, Y., Sun, J., 2020: Deep learning based point cloud registration: an overview. *Virtual Reality & Intelligent Hardware*, 2(3), 222–246.

Zhou, Q., Park, J., Koltun, V., 2016: Fast global registration. *Computer Vision - ECCV 2016 - 14th European Confer- ence*, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II, 766–782.

Zhou, Z., Zhao, C., Adolfsson, D., Su, S., Gao, Y., Duckett, T., Sun, L., 2021: NDT-Transformer: Large-Scale 3D Point Cloud Localisation using the Normal Distribution Transform Representation. *ArXiv*: abs/2103.12292.