

# Method for registration of 3D shapes without overlap for known 3D priors

P. Hu  and A. Munteanu

Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium

## ✉ Correspondence

P. Hu, Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium.

Email: phu@etrovub.be

In 3D registration of point clouds, the goal is to find an optimal transformation that aligns the input shapes, provided that they have some overlap. Existing methods suffer from performance degradation when the overlapping ratio between the neighbouring point clouds is small. So far, there is no existing method that can be adopted for aligning shapes with no overlap. In this letter, to the best of knowledge, the first method for the registration of 3D shapes without overlap, assuming that the shapes correspond to partial views of a known semi-rigid 3D prior is presented. The method is validated and compared to existing methods on FAUST, which is a known dataset used for human body reconstruction. Experimental results show that this approach can effectively align shapes without overlap. Compared to existing state-of-the-art methods, this approach avoids iterative optimization and is robust to outliers and inherent inaccuracies induced by an initial rough alignment of the shapes.

**Introduction:** 3D registration is a classical and fundamental problem for countless applications. Since commodity depth cameras become less expensive and more accurate, depth images play an increasingly important role in numerous tasks [1]. In order to obtain comprehensive information from 3D scenery, point clouds captured from multiple views need to be aligned. The well-established method is iterative closest point (ICP) [2] based on which a myriad of flavours have been proposed. In ICP, given a source shape and a target shape, the following steps are performed: (1) for each point in the source shape, identify the closest corresponding point in the target shape; (2) predict the transformation by minimizing the mean square Euclidean distance between these correspondences; (3) transform the source shape using the predicted transformation from step 2; (4) iterate the above steps until the mean square distance reaches a pre-defined threshold. ICP and its variants are the dominating methods for the task of 3D registration. However, ICP-based methods assume that the source and target shapes have been roughly aligned with a sufficient overlap..

Deep learning has shown its excellent ability to solve various problems which are difficult or impossible to address using traditional approaches. Recent research strives to explore 3D registration via deep learning [3], [4], [5], [6]. However, these methods are designed for shapes that partially overlap. In this letter, we present a novel deep learning method for 3D shape registration. Compared to the existing methods, the main advantage of our method is that we successfully handled the non-overlapping shape registration problem. We assume that the shapes correspond to partial views of a known semi-rigid 3D prior. This problem is impossible to be addressed using ICP due to the lack of point correspondences. This is addressed in this letter of which the main contributions can be summarized as follows:

- We propose a novel deep learning method for 3D registration; to the best of our knowledge, this is the first method for registration of 3D shapes without overlap.
- We present a novel learned correspondence representation.
- We validate the effectiveness of our approach by applying it to the challenging task of 3D human body reconstruction.

**Related work:** Methods for 3D shape registration can be classified into two main categories: (1) pairwise registration (e.g. [2]), and (2) multiview global registration (e.g. [7]) techniques. ICP and its variants are originally designed for pairwise registration, which takes two point clouds or meshes as input. Traditional pairwise registration needs two

steps: (1) the rough alignment, which provides the initial estimate of the relative transformation, and (2) to iterative refinement of the alignment by minimizing the 3D registration error. Pairwise registration does not perform well when operating on multi-view 3D scans due to the incremental pairwise registration errors. To address this problem, multi-view registration methods are proposed in order to refine the global registration by incorporating cues from multiple views [8, 9].

Recently, several researchers have proposed deep learning approaches for the task of 3D registration. [3] proposed deep closest point (DCP) that directly operates on point clouds. DCP consists of a point cloud feature extractor, an attention-based module and a differentiable singular value decomposition (SVD) layer to predict the rigid transformation. However, DCP followed by ICP is required to refine the registration results. PointNetLK [10] proposed the registration of point clouds by minimizing the distance between the learned global embedding resulting from PointNet [11]. PRNet [4] aligned point clouds with partial overlap based on DCP. These methods aim to jointly learn the feature vector and registration, which results in lack of generalization. DeepICP [12] learned correspondences between point clouds and then use ICP to apply SVD for registration. Similarly, 3DSmoothNet [13] use a Siamese deep learning architecture to establish correspondences between point clouds. [5] proposed an end-to-end solution for multiview 3D point cloud registration. However, these methods fail when there is no overlap between the inputs. In stark contrast to these methods, we aim at aligning point clouds or meshes without overlap, when knowing that they originate from an (semi-rigid) object with a given 3D shape. An example is to have the frontal and back view of the object. The two views do not overlap but they do originate from the same 3D object. Existing registration methods will fail due to lack of correct correspondences. To this end, we proposed a novel learned correspondence representation for inputs without overlap.

**Problem statement:** We use  $X$  and  $Y$  to denote the source and target point clouds respectively, where  $X = \{x_i \in \mathbb{R}^3, i = 1, \dots, N\}$  and  $Y = \{y_j \in \mathbb{R}^3, j = 1, \dots, M\}$ . The proposed method works for both  $N = M$  and  $N \neq M$ . This is a rigid registration problem, whereby we assume that  $X$  is transformed by an unknown rigid motion and then aligned with  $Y$ . We denote the rigid transformation as  $[R_{xy}, t_{xy}]$  where  $R_{xy} \in SO(3)$  and  $t_{xy} \in \mathbb{R}^3$ . Most of registration algorithms aim at minimizing the mean-squared error  $E(R_{xy}, t_{xy})$ , given by:

$$E(R_{xy}, t_{xy}) = \frac{1}{N} \sum_i ||R_{xy}x_i + t_{xy} - y_{\text{cor}}(x_i)||^2 \quad (1)$$

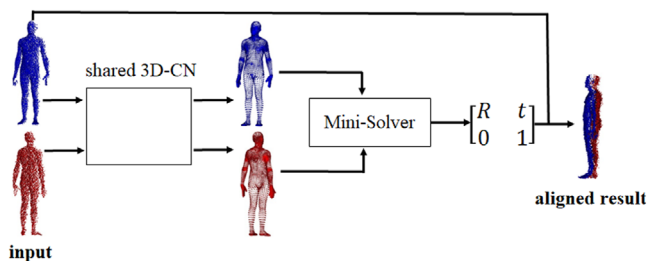
where  $\text{cor}()$  is an operation to find correspondences in  $Y$  for each point from  $X$ . Solving this optimization and finding correspondences are performed alternatively in an iterative manner. This formulation absolutely fails for our task due to lacking overlap or correspondences.

In contrast to existing formulation for registration, we propose to learn a novel correspondence representation  $\{\phi(X), \phi(Y)\}$  for  $X$  and  $Y$ , where  $\phi(X) = \{\phi(x_i) \in \mathbb{R}^3, i = 1, \dots, U\}$  and  $\phi(Y) = \{\phi(y_i) \in \mathbb{R}^3, i = 1, \dots, U\}$ . In this study, we train a deep neural network to learn  $\phi$ , which takes the partial point cloud as input and outputs the virtual correspondence. Then a perfect registration is obtained based on the correspondence representation. This can be written as:

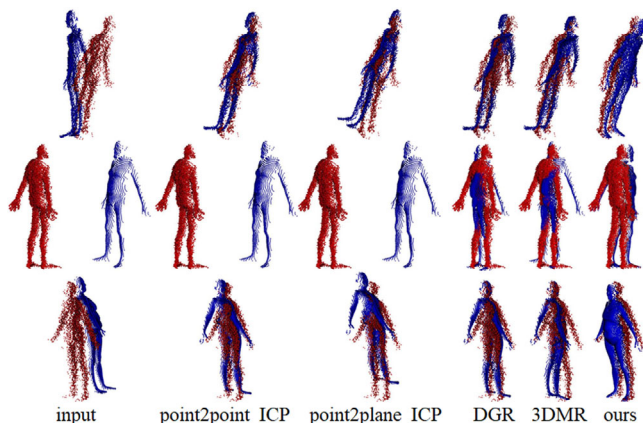
$$\begin{bmatrix} R_{xy} & t_{xy} \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} \phi(X) \\ \text{ones}(U) \end{bmatrix} = \begin{bmatrix} \phi(Y) \\ \text{ones}(U) \end{bmatrix} \quad (2)$$

$\text{ones}()$  creates a row vector filled with ones. Hence, the rigid transformation can be easily obtained using normal equation:

$$\begin{bmatrix} R_{xy} & t_{xy} \\ 0 & 1 \end{bmatrix} = \left( \begin{bmatrix} \phi(Y) \\ \text{ones}(U) \end{bmatrix} \times \begin{bmatrix} \phi(X) \\ \text{ones}(U) \end{bmatrix}^T \right) \times \left( \begin{bmatrix} \phi(X) \\ \text{ones}(U) \end{bmatrix} \times \begin{bmatrix} \phi(X) \\ \text{ones}(U) \end{bmatrix}^T \right)^{-1} \quad (3)$$



**Fig 1** Overview of the proposed approach. The source point cloud (in blue) and the target point cloud (in red) are fed into 3D-CN. No assumption is made about the overlap, rough alignment or the input order of points for the partial point clouds. The mini solver is used for computing the affine transformation between the two outputs produced by the network. Finally, the estimated transformation is applied to align the partial point clouds

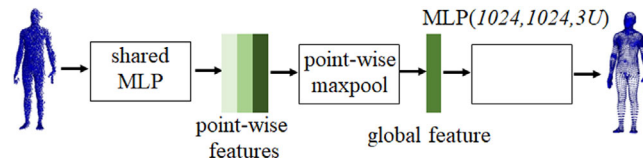


**Fig 2** Comparison with different registration methods based on FAUST data

Our method is an one-shot solution, and no iteration is needed.

**Proposed correspondence representation:** Traditional ICP and its variants utilize closest neighbours as correspondences. However, they are prone to the initialization and overlap ratio. PointNetLK [10] learned an embedding correspondence representation and implemented registration by minimizing the distance between the learned correspondences. This method, however, is prone to misalignment and not designed for non-overlap shape registration. In contrast to these methods, we propose to learn 3D correspondences for the task of shape registration without overlap. Given partial point clouds of the same subject, we train a neural network, dubbed 3D Correspondence Network (3D-CN), to output a fixed number (denoted by  $U$ ) of ordered 3D points that enables establishing the correspondence between the input partial point clouds and the complete geometric shape from which they originate.  $U$  is a hyperparameter in our model. This idea is inspired by point cloud completion [14]. In contrast to [14], the output produced by 3D-CN has the following properties: (1) it represents the same complete geometric shape (a complete body shape in the example of Figure 1), (2) it has the same number  $U$  of points for any input partial point cloud and (3) it has the same order of points, irrespective of the input. Based on these properties, one-to-one correspondences between partial point clouds of the same object are naturally built. Essentially, the outputs produced by 3D-CN are identical up to some affine transformation which can be found using a simple solver (see Figure 1). The complete object is a 3D prior that serves as proxy to enable the alignment between the partial point clouds. In the following ablation study, we perform experiments to analyse the effects of  $U$  on the registration.

**Proposed architecture:** Figure 1 shows the overview of the proposed method. Given two partial point clouds, a shared 3D Correspondence Network (3D-CN) is used to predict virtual correspondences for the source and target point clouds respectively. In this letter, we validate our algorithm by applying it to solve a challenging task: the human body



**Fig 3** Architecture of the proposed 3D Correspondence Network (3D-CN). Given a partial point cloud, we aim at predicting the virtual correspondence represented by the structured 3D point sets.  $U$  denotes the number of predicted correspondence points, which is can be controlled by the user

shape reconstruction from two non-overlapping scans. As 3D-CN takes one partial point cloud as input, our method can be directly used for multi-view point cloud registration. Next, the transformation parameters are directly obtained by a mini solver. As shown in Figure 3, 3D-CN mainly consists of two modules: a feature extractor from point clouds and a correspondence predictor. The feature extractor takes the point cloud  $X$  as input and extracts a  $k$ -dimensional feature vector  $f$  where  $k = 1024$ . Then, the correspondence predictor consumes the feature vector and output  $U$  structured 3D points. Our feature extractor is a simplified version of PointNet [11]. Its first layer takes  $m$  input points as input. A shared multi-layer perceptron (MLP) consisting of three linear layers with ReLU activation is designed to map each point to a point feature vector. Then, a point-wise max-pooling operation is used to obtain a global  $k$ -dimensional feature vector. Our correspondence predictor is designed by a MLP consisting of three linear layers with ReLU activation. The loss function  $L$  is defined by the mean squared error between the predicted correspondence  $\phi(X)$  and the ground truth correspondence  $\phi(X)^{GT}$ .

**Training dataset:** To demonstrate the effectiveness of our proposed approach, we train our model using a human body dataset. We generate  $1 \times 10^5$  synthetic human body shapes based on SMPL by sampling parameters from the SURREAL dataset [15], which is also used to train networks for other tasks [16]. The open-source Blender Sensor Simulation plugin Blensor [17] is used to render partial point clouds without overlap from the front-facing and back-facing views. We also set the parameter noise\_sigma=0.02 for adding noise to the point clouds when performing the experiments.

**Experiments:** We evaluate our results on the FAUST dataset by comparing the rotation and translation errors. We define the following metrics for quantitative comparison, which are also used in [6]. Given the estimated  $R$  and the ground truth  $R^{GT}$ , the rotation error is defined as:

$$RE(R, R^{GT}) = \arccos \left( \frac{\text{trace}(R^{-1}R^{GT}) - 1}{2} \right) \quad (4)$$

For the translation error, we use the following:

$$RE(t, t^{GT}) = \|t - t^{GT}\| \quad (5)$$

We compare our method against point-to-point ICP [2], point-to-plane ICP [18], deep global registration (DGR) [19] and 3D multiview registration (3DMR) [5]. The former two methods are popular ICP-based methods and the latter two are state-of-the-art deep learning-based registration approaches. Figure 2 provides results when given a source point cloud (in blue) and a target point cloud (in red) that share no overlap between them; results of different approaches are visually compared. It can be seen that ICP-based methods fail to perform the registration when the input two shapes are not well roughly aligned (Figure 2 middle row). Existing methods cannot obtain a correct registration when operating on inputs without overlap. However, experimental results show that our method can work well for the inputs without overlap even though the input data has noise and bad initial alignment. The quantitative comparisons of rotation and translation errors are reported in Table 1 and Table 2 respectively. The results show that our method significantly improves the registration accuracy compared to the existing state-of-the-art methods.

Table 1. Comparison of rotation errors with different registration methods

Method	point2point ICP	point2plane ICP	DGR	3DMR	ours
$\mu$	173.63	174.52	174.99	175.47	2.68
$\sigma$	2.21	4.20	5.73	3.50	3.79
max	176.57	178.32	179.58	179.47	8.05

Table 2. Comparison of translation errors with different registration methods

Method	point2point ICP	point2plane ICP	DGR	3DMR	ours
$\mu$	4.06	4.04	4.47	4.44	0.17
$\sigma$	0.11	0.10	0.56	0.58	0.03
max	4.20	4.18	5.25	5.26	0.21

Table 3. Rotation errors on different correspondences

Correspondence	4	10	100	1000	6890	CPC
$\mu$	12.45	5.43	5.58	5.63	5.62	169.49
$\sigma$	20.83	24.03	3.89	3.76	3.74	7.24
max	180.0	22.67	16.08	15.07	15.23	179.99

Table 4. Translation errors on different correspondences

Correspondence	4	10	100	1000	6890	CPC
$\mu$	0.04	0.00	0.00	0.00	0.00	0.13
$\sigma$	0.26	0.00	0.00	0.00	0.00	0.02
max	4.96	0.05	0.00	0.00	0.00	0.30

**Ablation study:** Based on the 400 testing data which is not included in the training, in this section we try to explore the effects on the registration of sparse and dense correspondences. We also compare our correspondences with the complete point clouds (CPC) from point completion networks [14]. Due to lacking the one-to-one correspondences for complete point clouds, point-to-point ICP is applied to compute the transformation. In this experiment, we manually down-sample 4, 10, 100, 1000 points from the SMPL body as the sparse correspondences. For a fair comparison, the number of output points from point completion networks is also set to 6890. As shown in Table 3 and Table 4, it can be observed that our method can significantly improve the registration accuracy compared to the point completion network. In addition, dense correspondences prove to yield more robust results compared to sparse correspondences.

**Conclusions:** To the best of our knowledge, this is the first method for 3D shape registration without overlap. We validate the effectiveness of our approach by applying it to the challenging task of 3D human body reconstruction from two partial non-overlapping scans. The results based on the FAUST dataset show that our approach yields robust results in non-overlap 3D registration. Our method proves also to be robust to noise and poor initial alignment, and it is not iterative. Comparisons with traditional registration methods based on iterative optimization as well as against recent deep learning registration approaches show that our method obtains state-of-the-art results by significantly improving the registration accuracy.

**Acknowledgement:** This work is supported by the eTailor project (Project ID BRGRD24)

© 2021 The Authors. *Electronics Letters* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Received: 1 November 2020 Accepted: 16 February 2021

doi: 10.1049/ell2.12133

## References

- Joukovsky, B., Hu, P., Munteanu, A.: Multi-modal deep network for rgb-d segmentation of clothes. *Electron. Lett.* **56**(9), 432–435 (2020)
- Besl, P.J., McKay, N.D.: Method for registration of 3-D shapes. In: *Sensor Fusion IV: Control Paradigms and Data Structures*, Vol. 1611, pp. 586–606. SPIE, Bellingham, WA (1992)
- Wang, Y., Solomon, J.M.: Deep closest point: Learning representations for point cloud registration. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3523–3532. IEEE, Piscataway, NJ (2019)
- Wang, Y., Solomon, J.M.: Prnet: Self-supervised learning for partial-to-partial registration. In: *Advances in Neural Information Processing Systems*, pp. 8814–8826. NeurIPS, Indianapolis, IN (2019)
- Gojcic, Z., et al.: Learning multiview 3d point cloud registration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1759–1769. IEEE, Piscataway, NJ (2020)
- Pais, G.D., et al.: 3dregnet: A deep neural network for 3d point registration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7193–7203. IEEE, Piscataway, NJ (2020)
- Pulli, K.: Multiview registration for large data sets. In: *Second International Conference on 3-D Digital Imaging and Modeling*, pp. 160–168. IEEE, Piscataway, NJ (1999)
- Sharp, G.C., Lee, S.W., Wehe, D.K.: Multiview registration of 3d scenes by minimizing error between coordinate frames. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(8), 1037–1050 (2004)
- Birdal, T., Ilic, S.: Cad priors for accurate and flexible instance reconstruction. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 133–142. IEEE, Piscataway, NJ (2017)
- Aoki, Y., et al.: Pointnetlk: Robust & efficient point cloud registration using pointnet. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7163–7172. IEEE, Piscataway, NJ (2019)
- Qi, C.R., et al.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660. IEEE, Piscataway, NJ (2017)
- Lu, W., et al.: Deepicp: An end-to-end deep neural network for 3d point cloud registration. arXiv:1905.04153 (2019)
- Gojcic, Z., et al.: The perfect match: 3d point cloud matching with smoothed densities. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5545–5554. IEEE, Piscataway, NJ (2019)
- Wen, X., et al.: Point cloud completion by skip-attention network with hierarchical folding. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1939–1948. IEEE, Piscataway, NJ (2020)
- Varol, G., et al.: Learning from synthetic humans. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 109–117. IEEE, Piscataway, NJ (2017)
- Hu, P., et al.: Learning to estimate the body shape under clothing from a single 3d scan. *IEEE Transactions on Industrial Informatics* (2020) DOI: <https://doi.org/10.1109/TII.2020.3016591>
- Gschwandtner, M., et al.: Blensor: Blender sensor simulation toolbox. In: *International Symposium on Visual Computing*, pp. 199–208. Springer, Berlin (2011)
- Park, S.-Y., Subbarao, M.: An accurate and fast point-to-plane registration technique. *Pattern Recognit. Lett.* **24**(16), 2967–2976 (2003)
- Choy, C., Dong, W., Koltun, V.: Deep global registration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2514–2523. IEEE, Piscataway, NJ (2020)