Scientific Research Publishing

Journal of Computer and Communications, 2018, 6, 116-125
http://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Point Reg Net: Invariant Features for Point Cloud Registration Using in Image-Guided Radiation Therapy

Zhengfei Ma, Bo Liu*, Fugen Zhou, Jingheng Chen

Image Processing Center, Beihang University, Beijing, China
Email: *bo.liu@buaa.edu.cn

How to cite this paper: Ma, Z.F., Liu, B., Zhou, F.G. and Chen, J.H. (2018) Point Reg Net: Invariant Features for Point Cloud Registration Using in Image-Guided Radiation Therapy. *Journal of Computer and Communications*, **6**, 116-125.
https://doi.org/10.4236/jcc.2018.611011

Received: September 25, 2018
Accepted: November 12, 2018
Published: November 19, 2018

Abstract

In image-guided radiation therapy, extracting features from medical point cloud is the key technique for multimodality registration. This novel framework, denoted Control Point Net (CPN), provides an alternative to the common applications of manually designed key-point descriptors for coarse point cloud registration. The CPN directly consumes a point cloud, divides it into equally spaced 3D voxels and transforms the points within each voxel into a unified feature representation through voxel feature encoding (VFE) layer. Then all volumetric representations are aggregated by Weighted Extraction Layer which selectively extracts features and synthesize into global descriptors and coordinates of control points. Utilizing global descriptors instead of local features allows the available geometrical data to be better exploited to improve the robustness and precision. Specifically, CPN unifies feature extraction and clustering into a single network, omitting time-consuming feature matching procedure. The algorithm is tested on point cloud datasets generated from CT images. Experiments and comparisons with the state-of-the-art descriptors demonstrate that CPN is highly discriminative, efficient, and robust to noise and density changes.

Keywords

Medical Image Registration, Point Cloud, Deep Learning, Invariant Feature

1. Introduction

1.1. Overview

A point cloud is a set of unorganized irregular 3D points in a unified coordinate system, capturing 3D spatial information of an object or scenery. 3D point cloud

registration refers to a research problem in which two isomorphic point cloud sets are known by coordinates respectively, and an optimal homomorphous transformation between them is remained to be solved. It is a fundamental problem in 3D computer vision. Many applications have been developed including 3D modeling [1] [2], object recognition [3], surface alignment [4] [5] and antonymous driving [6].

There is no overwhelming solution for this challenging task due to unknown initial positions, noisy raw data, varying point densities, partial loss or overlap, and more unknown difficulties. Generally, current point cloud registration methods can be classified into two categories: iteration-based [7] and feature-based [8] [9] [10]. Iteration-based methods inevitably carry heavy computational load which is unwanted in the scenery of real-time surgery and image-guided radiation therapy.

In this paper, we focus on the feature-based method in which registration is achieved by extracting features, matching features and generating point-to-point correspondences, which is independent of initial proximity information between coordinate systems. For feature-based methods, local feature descriptors play a core role in feature matching [11]. In general, a good feature descriptor should be highly descriptive, unambiguous, compact and computationally efficient to enable a good matching performance. And the descriptors should also be robust to common noises and adaptive to various modalities, which is a challenging task due to the varieties of nuisances and the point cloud's irregular format.

At present, numerous local feature descriptors have been demonstrated [9] [12] [13] [14]. All of these feature descriptors either make use of local geometric statistics (e.g., normal, curvature), or spatial distribution of the neighboring points. For almost planar or spherical surfaces, they either suffer from low descriptiveness, or sensitiveness for surface noise and single outliers, as a result of lack of global topology information. In addition, descriptors derived from mathematical formula doubtfully make fully use of common statistical characteristics of a given dataset.

In this context, inspired by [15], we conduct a novel type of deep learning network that extracts feature descriptors aiming at dataset of a specific category, which makes it possible to take advantage of characteristics of medical data and its unique statistics distribution. Softmax layer and Max Pooling layer are innovatively modified to be Weighted Extraction Layer (WEL), and are applied to point-wise information. Local features are weighted, synthesized and aggregated into global descriptors as well as coordinates of control points, which are mean coordinates across local subsets of points. Utilizing control points instead of key-points allows the available geometrical data to be better exploited [16], and improves the robustness of registration. The global descriptors and coordinates of control points are internally promising sorted, omitting the time-consuming and error-prone feature matching procedure.

The contributions of this work are as follows:

1) We improve the architecture of voxel feature encoding (VFE) layer based on Voxel Net [6], which is suitable for consuming unordered point sets in 3D;

2) We propose Weighted Extraction Layer (WEL) which selectively synthesizes local features into global descriptors and coordinates of control points;

3) We show how CPN can be trained to perform 3D registration without point matching procedure;

1.2. Related Works

Registration of pre- and intra-interventional data is one of the enabling technologies for image-guided radiation therapy, radio surgery, and interventional radiology. Different combinations of modalities and dimensionalities, either 3D/2D or 3D/3D, have been studied [17] [18]. However, seldom has application adopted point cloud as format of input data. [19] presents an 2-D/3-D registration based on CNN regression using for image-guided interventions. [16] explores the use of 3D point cloud sensors in medical augmented reality applications.

On the other side, many studies on general point cloud registration have been conducted, e.g. spin images [12], fast point feature histograms (FPFH) [9], signature of histograms of orientations (SHOT) [13], and rotational projection statistics (RoPS) [14]. [20] proposes local feature statistics histogram (LFSH)feature descriptor, and gives a performance comparison between these features.

Recent days, point cloud processing based on deep learning has become research focus. [21] provides a classic unified architecture for unordered point cloud classification, segmentation, and semantic parsing. [6] unifies feature extraction and bounding box prediction into a single end-to-end trainable deep network. [15] represents latest progress of deep-learning-based algorithm for point cloud registration. It encodes local 3D geometric structures into super-points using unsupervised auto-encoder. However, a matching procedure and a fine-tuning stage must be performed before transformation estimation.

2. Algorithm Principle

The CPN takes as input two point clouds which are similar in shape by different in local coordinates, extracting local features per voxel, then synthesizes them into coordinates of control points. One group of control points paired with one point cloud, representing its position and pose. The global descriptors and coordinates of control points are internally promising sorted, therefore transformation estimation with RANSAC can be immediately applied without feature matching procedure. The number of control points M is fixed regardless of point cloud size. Currently we set $M = 64$.

2.1. Control Point Net Architecture

The proposed CPN consists the following layers or functional blocks:

1) Data Grouping, in which the 3D space is equally subdivided into cubic voxels, points are grouped according to the voxel they reside in;

2) Random Sampling, in which fixed number inside each voxel are randomly selected;

3) Voxel Feature Encoding (VFE) Layers, in which point-wise features and locally aggregated feature are combined to learn descriptive shape;

4) Multiple Fully Connected Layers, in which each point-wise feature are aggregated within a middle layer, then packed up into a fixed-length feature vector;

5) Weighted Extraction Layer, in which feature vectors in a same voxel are first normalized by softmax arithmetic along elements, then synthesized to be the feature vector of this voxel. Same operations apply to coordinate. Finally, features and coordinates extracted from all voxels are maxpooled along elements and the global descriptors and super-point coordinates are generated. The architecture of CPN is illustrated in **Figure 1**.

2.2. Data Grouping and VFE

These blocks are designed based on operations in [6], aiming to extract and learn descriptive shape information from local point distribution. Points are subdivided and grouped along Z, Y, X respectively into voxels. Points with highly variable density throughout different voxels are randomly sampled to keep a fixed number T.

As shown in **Figure 2**, the architecture of VFE Layer consists of a point-wise shared fully connected layer, an element-wise maxpooling layer, and a doubled point-wise concatenated feature vector as output.

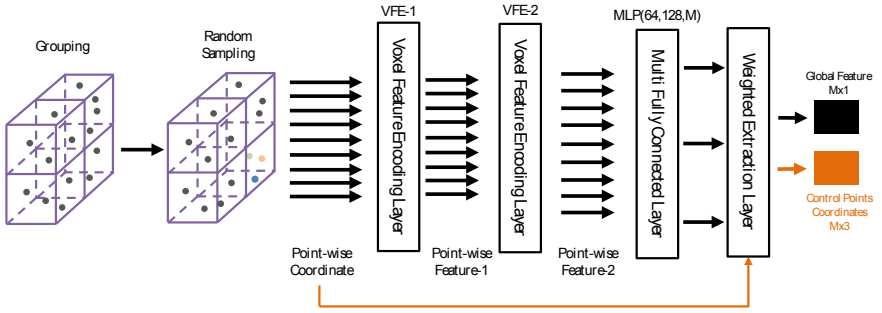

Figure 1. Architecture of CPN.

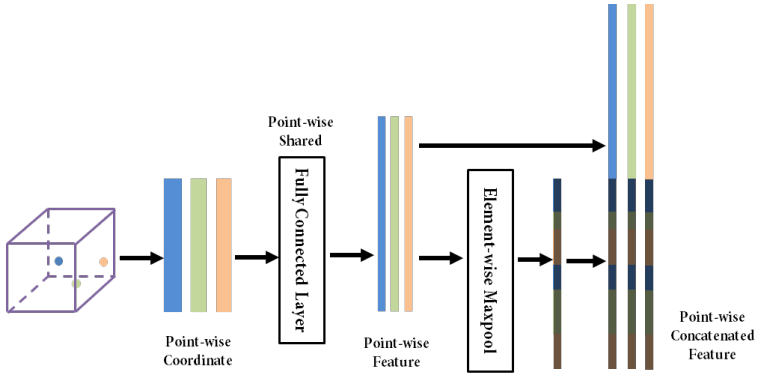

Figure 2. Voxel feature encoding layer.

The inputs of VFE-1 is

$$V = \left\{ \mathbf{p}_i = \left[x_i, y_i, z_i, x_i - c_x, y_i - c_y, z_i - c_z \right] \in \Re^6 \right\}$$

where the centroid of points in each voxel is denoted as $\left(c_x, c_y, c_z \right)$.

The inputs of VFE-2 are the outputs of VFE-1.

2.3. Multiple Fully Connected Layers

We design a 3-layer fully connected network, with Batch Normalization and ReLU applied sequentially. All layers are operated on point-wise feature generated from VFE-2 layer. Layers share same weights for all points.

2.4. Weighted Extraction Layer

In order to weight and synthesize point-wise information, softmax arithmetic on point-wise features and coordinates in same voxel is first applied.

Let $F_{kti} = \left\{ \left\{ f_i \right\}_{1\ldots T} \right\}_{1\ldots K}$ be the inputs of softmax block, where T is the number of points in a voxel, K is the number of non-empty voxels. $\left\{ f_i \right\}_{i=1\ldots M}$ is the feature vector generated from Fully Connected Layers, where M is output dimension.

Let $\mathbf{P}_{kt} = \left\{ \left\{ (x, y, z) \right\}_{1\ldots T} \right\}_{1\ldots K}$ be the coordinates of points in point cloud.

$$\widetilde{Q}_{ki} = \sum_{t=1}^{T} \underset{t}{SoftMax}\left(F_{kti} \right) \bullet F_{kti} \tag{1}$$

$$\widetilde{\mathbf{P}}_{ki} = \sum_{t=1}^{T} \underset{t}{SoftMax}\left(F_{kti} \right) \bullet \left(\mathbf{P}_{kt} \right) \tag{2}$$

are the voxel-wise softmax feature vectors and coordinate vectors, where the definition of softmax arithmetic is

$$\underset{t}{SoftMax}\left(X_t \right) = \frac{e^{X_t}}{\sum_t e^{X_t}} \tag{3}$$

Then feature vectors and coordinate vectors are extracted from all voxels are maxpooled into global descriptors and control points

$$\widehat{Q}_i = \underset{k}{Max}\, \widetilde{Q}_{ki} \tag{4}$$

$$\widehat{\mathbf{P}}_i = \widetilde{\mathbf{P}}_{k'i} \tag{5}$$

where k' in (5) is paired with k in (4) (**Figure 3**).

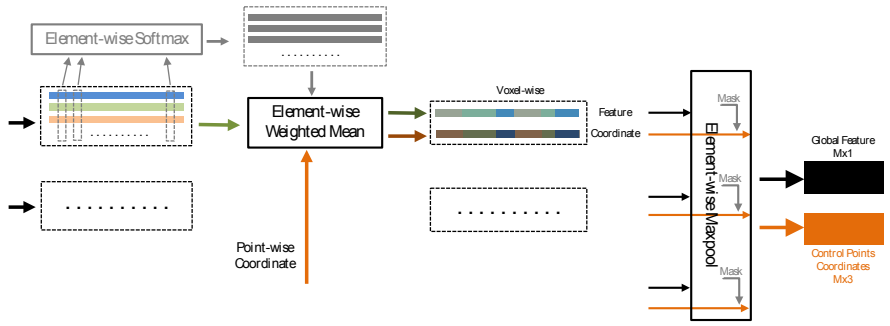

Figure 3. Weighted extraction layer.

2.5. Loss Function

CPN is designed to extract features for registration. In order to achieve minimum registration error, the coordinates of control points should be invariant under the point cloud coordinate system. Thus for every sample in dataset, we define loss function as

$$loss = \sum_i \left\| \hat{\mathbf{P}}_{Ai} - \mathbf{H} \cdot \hat{\mathbf{P}}_{Bi} \right\| \tag{5}$$

where $\mathbf{H}$ is rigid transformation matrix that projects corresponding points from one point cloud onto another. $\left\{\hat{\mathbf{P}}_i\right\}$ are coordinates of control points generated from CPN. The L1-Norm is preferred to utilize robustness upon existence of outlier points.

3. Experimental

3.1. Dataset

We're mostly interested in the application on medical scenery such as real-time surgery or image-guided radiation therapy. Thus we establish a medical point cloud dataset using for both training and evaluating. The raw data are download from The Cancer Imaging Archive (TCIA) [22]. Collections of CT scans for lung cancer (National Lung Screening Trial, LCTSC, NSCLC-Radiomics) are adopted. All point clouds are extracted from CT images by edge detection along Z-axis positive direction. The intensity threshold is set to −250, which effectively detects the surface of skin. On average, 40,000 points can be extracted from a series of CT images and aggregated into a point cloud sample. Samples that have too few points are dropped. Samples that have too many points are down-sampled. All samples are cropped into a same 3D size, 40 cm × 40 cm × 40 cm, and the mass centers are translated into origin of coordinates. In total, we prepared 1000 samples for training and 100 for evaluating.

Labels are homograph transforms $\mathbf{H}$ whose rotation part is generated as a uniform distribution in $\Re^3$, and translation part is a uniform distribution between (−5 cm, 5 cm). Every training sample is composed of two identical point clouds, and can be transformed from one to another by homograph $\mathbf{H}$ (**Figure 4**).

With only 1000 training sample, training our network from scratch will undoubtedly suffer from overfitting. We introduce three kinds of data augmentation techniques to relieve this tendency, including jitter on points, scaling on points, and jitter on transform $\mathbf{H}$. The augmentations are all done on-the-fly by GPU without being stored on disk.

3.2. Experimental Setup

All experiments are conducted under Intel Corel i7-5960X, 32.0GB RAM, NVIDIA GeForce GTX1080Ti, Windows 7 SP1. The development environment is Tensorflow 1.2.0, Python 3.6.0.

The voxel size is a tradeoff parameter between precision and speed. It's also a

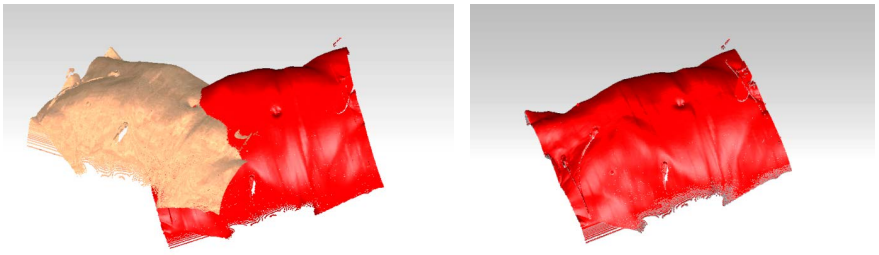

Figure 4. Dataset.

bound between local details and global coverage. Currently we adopt 10 × 10 × 10 as the size of grid.

3.3. Evaluation and Analysis

For quantitative analysis of the correspondences, we use percentage of success correspondences [22]. When the transformation error $e = \left\| \hat{\mathbf{P}}_{Ai} - \mathbf{H} \cdot \hat{\mathbf{P}}_{Bi} \right\|$ is below a predefined threshold (we used 1 cm), the registration result is marked "success". The smaller the RMSE, the better the two point clouds are aligned. The output of CPN can be used to calculate success correspondences directly. In contrast, for classic feature extractors, a matching algorithm should be first applied to obtain feasible correspondences between $\hat{\mathbf{P}}_A$ and $\hat{\mathbf{P}}_B$. We evaluated the performance on testing dataset with Gaussian noise (standard deviation = 1 mm). The comparison results are shown in **Table 1**. Parameters for classic feature descriptors are set as **Table 2**.

Table 1 suggests that CPN outperforms all classic descriptors in the percentage of success correspondences (71.4%) and achieves a promising matching result.

The time efficiency comparison is illustrated in **Table 3.** All operations in CPN except the data grouping and sampling blocks are regular and dense, enabling the acceleration by GPU in a parallel manner. One can see that our CPN is superior to other descriptors in time efficiency. FPFH and SHOT descriptors are faster than the PFH descriptor by approximately an order of magnitude. In addition, the time for RANSAC matching should not be neglected.

To verify the effectiveness and robustness of the registration algorithm under nuisances, we apply modifications on test data including: randomly moving 5% - 30% of the points a uniformly distributed distance of up to 5 mm; randomly removing 5% - 30% of the points to test sensitivity to the density change.

Figure 5(a) and **Figure 5(b)** summarize the results. All descriptors can keep a relative high percentage of success correspondences under low-level noise or tiny change of density. The CPN always performs better compared with classic descriptors. However, the performance of CPN drops quickly when more points are removed. Even in challenging case with 30% points removed, our network is still workable and gives an acceptable percentage of success.

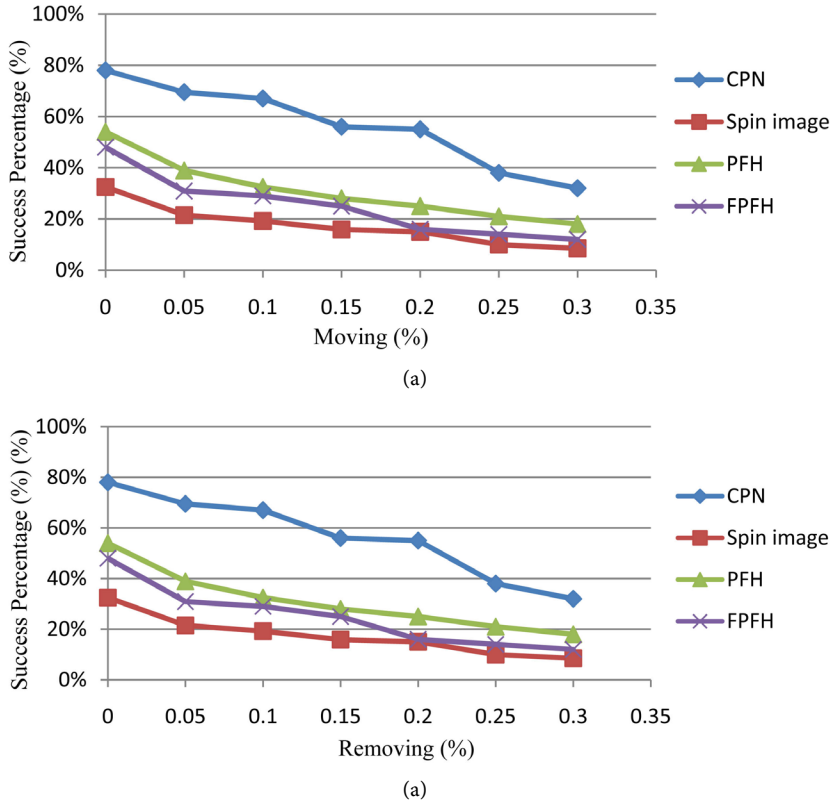

Figure 5. (a) Success percentage under moving; (b) Success percentage under removing.

Table 1. Performance comparison (Averaged).

Criteria	CPN	Spin image	PFH	FPFH	SHOT
NC	64	451	295	210	410
PSC	**71.4%**	23.1%	28.4%	19.7%	35.5%

a. NC denotes the number of correspondences. For CPN, all super-point coordinates are regarded as correspondences; b. PSC represents the percentage of success correspondences.

Table 2. Parameter settings.

	Spin image	PFH	FPFH	SHOT
Search radius (%)	-	13	13	13
Support angle (degree)	60	-	-	-
Dimensionality	15 × 15	5 × 5 × 5	3x11	8 × 2 × 2 × 10
length	225	125	33	320

Table 3. Time efficiency (Averaged).

Criteria	CPN	Spin image	PFH	FPFH	SHOT
Extraction	1.7s	83s	280s	14s	19s
Matching		26s	14s	6s	22s

a. CPN is accelerated by GPU. Classic features and RANSAC matching run on single CPU.

4. Conclusions

This paper presented CPN, an innovative network that extracts feature descriptors and control points for medical point cloud registration. The CPN deals with the challenges of extracting robust and unambiguous point cloud features and achieve great improvement over state-of-art algorithms.

A possible explanation why CPN outperforms classic feature descriptors is that the training stage makes it possible for network to utilize a certain assumption on input data statistical characteristics and distribution.

A possible explanation why CPN is much faster than classic feature descriptors is that the architecture of CPN is easily accelerated by parallel computational architecture. While it includes an offline training stage, the online stages can be implemented efficiently in parallel, making it suitable for serving real-time applications

In future work, we intend to adapt this approach for scalable size of voxel. Another interesting direction is to design a multi-scale version of CPN, similarly as Point Net++ [23].

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61601012).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[2] Lu, K., Wang, Q., Xue, J. and Pan, W. (2014) 3d Model Retrieval and Classification by Semi-Supervised Learning with Content-Based Similarity. *Information Sciences: An International Journal*, **281**, 703-713. https://doi.org/10.1016/j.ins.2014.03.079

[3] Guo, Y., Sohel, F., Bennamoun, M., Wan, J. and Lu, M. (2015) A Novel Local Surface Feature for 3d Object Recognition under Clutter and Occlusion. *Information Sciences*, **293**, 196-213. https://doi.org/10.1016/j.ins.2014.09.015

[4] Albarelli, A., Rodolà, E. and Torsello, A. (2015) Fast and Accurate Surface Alignment through an Isometry-Enforcing Game. *Pattern Recognition*, **48**, 2209-2226. https://doi.org/10.1016/j.patcog.2015.01.020

[5] Gong, M., Wu, Y., Cai, Q., Ma, W., Qin, A.K., Wang, Z., *et al*. (2016) Discrete Particle Swarm Optimization for High-Order Graph Matching. *Information Sciences: An International Journal*, **328**, 158-171. https://doi.org/10.1016/j.ins.2015.08.038

[6] Zhou, Y. and Tuzel, O. (2017) Voxelnet: End-to-End Learning for Point Cloud Based 3d Object Detection.

[7] Chen, C.S., Hung, Y.P. and Cheng, J.B. (2002) Ransac-Based Darces: A New Approach to Fast Automatic Registration of Partially Overlapping Range Images. *IEEE*

Transactions on Pattern Analysis & Machine Intelligence, **21**, 1229-1234. https://doi.org/10.1109/34.809117

[8] Guo, Y., Sohel, F., Bennamoun, M., Wan, J. and Lu, M. (2014) An Accurate and Robust Range Image Registration Algorithm for 3d Object Modeling. *IEEE Transactions on Multimedia*, **16**, 1377-1390. https://doi.org/10.1109/TMM.2014.2316145

[9] Rusu, R.B., Blodow, N. and Beetz, M. (2009) Fast Point Feature Histograms (FPFH) for 3D Registration. *IEEE International Conference on Robotics and Automation*, IEEE Press, 1848-1853. https://doi.org/10.1109/ROBOT.2009.5152473

[10] Wyngaerd, J.V., Koch, R., Proesmans, M. and Gool, L.V. (1999) Invariant-Based Registration of Surface Patches. *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, **1**, 301-306. https://doi.org/10.1109/ICCV.1999.791234

[11] Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J. and Kwok, N.M. (2016) A Comprehensive Performance Evaluation of 3d Local Feature Descriptors. *International Journal of Computer Vision*, **116**, 66-89. https://doi.org/10.1007/s11263-015-0824-y

[12] Johnson, A.E. and Hebert, M. (1998) Surface Matching for Object Recognition in Complex Three-Dimensional Scenes. *Image & Vision Computing*, **16**, 635-651. https://doi.org/10.1016/S0262-8856(98)00074-2

[13] Tombari, F., Salti, S. and Stefano, L.D. (2010) Unique Signatures of Histograms for Local Surface Description. *European Conference on Computer Vision Conference on Computer Vision*, **6313**, 356-369. https://doi.org/10.1007/978-3-642-15558-1_26

[14] Lu, M., Wan, J.W., Guo, Y.L., *et al*. (2013) Rotational Projection Statistics for 3d Local Surface Description and Object Recognition. *International Journal of Computer Vision*, **105**, 63-86. https://doi.org/10.1007/s11263-013-0627-y

[15] Elbaz, G., Avraham, T. and Fischer, A. (2017) 3D Point Cloud Registration for Localization Using a Deep Neural Network Auto-Encoder. *Computer Vision and Pattern Recognition*, 2472-2481.

[16] Placitelli, A.P. and Gallo, L. (2011) 3D Point Cloud Sensors for Low-Cost Medical *In-Situ* Visualization. *IEEE International Conference on Bioinformatics and Biomedicine Workshops*, 596-597. https://doi.org/10.1109/BIBMW.2011.6112435

[17] Markelj, P., Tomaževič, D., Likar, B. and Pernuš, F. (2012) A Review of 3d/2d Registration Methods for Image-Guided Interventions. *Medical Image Analysis*, **16**, 642-661. https://doi.org/10.1016/j.media.2010.03.005

[18] Shams, R., Sadeghi, P., Kennedy, R.A. and Hartley, R.I. (2010) A Survey of Medical Image Registration on Multicore and the Gpu. *Signal Processing Magazine IEEE*, **27**, 50-60. https://doi.org/10.1109/MSP.2009.935387

[19] Miao, S., Wang, Z.J. and Liao, R. (2016) A Cnn Regression Approach for Real-Time 2d/3d Registration. *IEEE Transactions on Medical Imaging*, **35**, 1352-1363. https://doi.org/10.1109/TMI.2016.2521800

[20] Yang, J., Cao, Z. and Zhang, Q. (2016) A Fast and Robust Local Descriptor for 3d Point Cloud Registration. *Information Sciences*, **346-347**, 163-179. https://doi.org/10.1016/j.ins.2016.01.095

[21] Charles, R.Q., Su, H., Mo, K. and Guibas, L.J. (2017) PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, 77-85. https://doi.org/10.1109/CVPR.2017.16

[22] The Cancer Imaging Archive. http://www.cancerimagingarchive.net/

[23] Qi, C.R., Yi, L., Su, H. and Guibas, L.J. (2017) Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space.