Signal Processing 217 (2024) 109309

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Robust registration and learning using multi-radii spherical polar Fourier transform

Alam Abbas Syed [a,*], Hassan Foroosh [b]

[a] Department of Electrical and Computer Engineering, 4000 Central Florida Blvd, Orlando, 32816, FL, USA
[b] Department of Computer Science, 4000 Central Florida Blvd, Orlando, 32816, FL, USA

ARTICLE INFO

Dataset link: https://github.com/syedalamabbas/Robust-Registration-Learning-SPFT

Keywords:
Registration
Classification
Rotation estimation
Spherical convolutional networks
Spherical polar Fourier transform
Non-uniform fast Fourier transform
Spherical correlations
Spherical convolutions
Spherical harmonics
Phase correlation
Sub-pixel registration
Sub-voxel registration

ABSTRACT

This paper presents effective methods using spherical polar Fourier transform data for two different applications, with active areas of research, one as a conventional volumetric registration algorithm and other as a machine learning classification network. For registration purposes the proposed method has the following advantageous features: (i) it is a unique and effective technique for estimating up to 7 degrees of freedom for 3D volumetric registration, that has a closed-form solution for 3D rotation estimation, and which does not resort to recurrence relations or search for point correspondences between two objects/volumes, (ii) it allows for robust rotation estimation determined simultaneously on multiple spectral spheres, therefore complete stack of such spherical layers can be processed concurrently to obtain accurate and optimal all three angles, and (iii) it has the ability to handle arbitrary large rotation angles, is shown to be robust against the presence of noise, holes/missing data, and partial overlaps. We demonstrate the effectiveness of our solution with extensive experimentation, including a set of scanned MRI images, a crashed car parking dataset, and the Princeton shape benchmark dataset with hundreds of 3D objects. For the classification solution we modify and adapt an existing network in the literature, a type of spherical convolutional network, that is suitable for processing multi-radii spectral spherical data, and showcase the resulting robustness achieved in classification of objects from the ModelNet40 dataset, especially in the presence of outliers, additive noise, and missing data.

1. Introduction

Volumetric registration is the process of spatially aligning two or more 3D datasets of an object or a scene. The goal is to estimate the most optimal geometric transformation between 3D point clouds or volumes acquired by different sensors or from different viewpoints, with the goal of enabling the fusion of information and/or the estimation of the motion of sensors.

Registration is a widely studied problem that still remains an important topic of research. The algorithms proposed in the literature could be classified into *spatial domain* and *frequency domain* methods depending on whether the registration algorithm is operated using respective domain information. Our paper specifically contributes to a research area on the intersection of Fourier-domain methods and 3D volumetric registration. Registration using Fourier-domain information has received extensive attention due to its robustness and resilience to frequency-dependent and correlated noise, with a diverse range of applications including in remote sensing [1–5], medical imaging [6–8], biometrics [9–11], underwater sonar [12,13], and seismology [14,15], to name a few. An excellent review of the application areas can be found in [16].

3D registration is also a fundamental problem in computer vision, robotics, and graphics that is crucial for related problems such as object pose estimation, shape completion, shape retrieval, grasping/manipulation, identification/recognition, self-localization, robot kinematics, flight simulation, and even motion capture for animations. Convolutional Neural Networks (CNNs) have proven to be very useful and powerful in analyzing 2D planar images due to their invariance to local deformations, and their equivariance to translations in Euclidean spaces. However, they are not rotationally equivariant, and global rotational equivariance was sought through the use of large data augmentation that usually results in increased model complexity, poor accuracy and performance [17], and yet patch-wise rotational equivariance is not achieved. Worral et al. [17] designed a CNN that uses filters derived from complex harmonics in the 2D polar domain to achieve both rotational and translation equivariance simultaneously. In 3D, or in general higher dimensions, this rotational nuisance becomes much more pronounced and is not trivially addressable effectively

* Corresponding author.
E-mail addresses: syedalamabbas@gmail.com (A.A. Syed), Hassan.Foroosh@ucf.edu (H. Foroosh).
URLs: https://github.com/syedalamabbas (A.A. Syed), https://www.cs.ucf.edu/person/hassan-foroosh/ (H. Foroosh).

https://doi.org/10.1016/j.sigpro.2023.109309
Received 19 May 2023; Received in revised form 27 October 2023; Accepted 27 October 2023
Available online 4 November 2023
0165-1684/© 2023 Elsevier B.V. All rights reserved.

using data augmentation, see [18,19]. In the past few years, there is an emergence of a promising new class of group equivariant neural networks in the machine learning literature, that leverages the concepts of group representation theory and harmonic analysis, where the designed networks have turned out to be of low model complexity and can still perform well in the presence of rotational nuisance in the input data because of their special structure [18–20]. Unlike regular Euclidean domain CNNs, these group equivariant special CNNs represent the 3D data on sphere(s), and hence use signal processing tools of the spherical domain, such as spherical convolution/correlation and spherical harmonics to achieve state of the art results in shape classification, retrieval and alignment, in the presence of rotational nuisance and other forms of perturbation in the input data. A number of problems of recent interest have created a demand for solutions that can analyze 3D spherical images like omnidirectional vision for drones, robots, and automated cars, molecular regression, global weather and climate modeling [19,21].

Fourier transform (FT) or spectral information acts as a global, featureless and robust descriptor [22] of a signal with the following advantages over the spatial algorithms:

- **Robustness**: Known stability to additive noises, distortions, spatial/time varying illumination disturbances, as they use all the information available rather than the content associated with only a sets of features.
- **Reliability**: Less sensitive to any specific control parameters, determined only by their run time and precision.
- **Separability**: Due to the nature of the FT a full separability of affine matrix and the translation parameters of a complete affine transformation is achieved.
- **Speed**: Low computational complexity, require no initialization, and take a fixed time period for registration.
- **Versatility**: The fundamental phase correlation method [23,24] for image translation registration can be easily extended to sub-pixel/sub-voxel accuracy [25–28], and even in the case of a partial overlap [29].

For a Euclidean 3D transformation between the two volumes/objects, the problem of estimating the rotation between two 3D objects is still under active research since issues such as missing data, partial overlap, noisy data, and objects with intraclass variations/symmetries have not been completely solved [30]. Some techniques are better suited to deal with only surface data [30], while others are designed for a full volumetric data [31]. In this paper, we focus on the volume data as surface data can be considered a subset of it. In 3D robotic mapping context of directly using 3D points, the well known Iterated Closest Point (ICP) algorithm [32] is most popular, see also [33,34] for the recent development. But such spatial registration approaches have known disadvantages as discussed in the literature [30,31], see also [35,36] for details. In general, methods that take the global appearances of 3D scans, such as Fourier based methods, see [29–31,37–40], are considered to be more robust especially in the presence of significant noise and other distortions.

Our algorithm is inspired by Frequency-based approaches present in [30,31,39,40]. In the previous work of Lucchese et al. [39]–Keller et al. [41], an angle-axis representation for 3D rotation is used. A critical point is the determination of this rotation axis in cases where the scan data has interference and occlusions. The method followed in [30] is a robust quaternion based eigenvector decomposition solution using spherical harmonics (SH). It works by incorporating a pre-processing stage in which the holes are filled based on a user parameter selection. This process requires an accumulation of all the data points for the volume onto a single unit sphere in order to evaluate the rotation angles using SH. As discussed further in the Section 3.1, the fine volumetric or point set structural information is largely lost in the process of this accumulation, hence most of these methods lose accuracy in the presence of noise, distortions, holes/missing data and partial overlaps. All the methods in [30,39,41] use the integration/accumulation of Fourier magnitude data to determine the angle-axis rotation, which becomes problematic in the presence of holes/missing data or noise. Only the technique presented in [31] is a similar multi-layered spectral approach, compared to our proposed method, with no accumulation or integration stage of the magnitude data for the 3D rotation estimation. However, it is based on interpolation to achieve resampling of Fourier domain information to spherical coordinates, which is one of the reasons why it can only estimate small rotation angles-detailed differences with our method and its advantages are discussed in the Section 3.1.

Rotation or rotational symmetries, naturally guide us towards representing the data on a simplest manifold: the sphere. Then the obtained spherical signals can be analyzed using the geometrical properties within this special domain such as SH. Such spherical representations for 3D data have been used in the literature well before the deep learning era due to their desirable properties in representation and analysis of rotational symmetries, see [42,43]. Indeed, the set of rotations on a sphere form a closed group of 3D manifold called as $\mathbf{SO}(\mathbf{3})$, the "special orthogonal group", and there are fast algorithms to compute spherical convolutions and correlations using SH. The principle of calculating a discrete Fourier transform (DFT) on a non-Cartesian grid was extended to 3D in [43] called the pseudo-polar (PP) grid and applied for 3D rotation estimation in [41]. In our paper we use spherical polar FT (SPFT) that we proposed in [44], which has the unique advantage of computing equiangular data points in the Fourier domain unlike the PP grid, see [44] for details. Before this there existed no such tool to obtain SPFT in the discrete setting, due to the prohibitive computational complexity. An analytical exact solution for estimating rotation angles from two equispaced spherical images was proposed in [45]. It uses SH on a rotation group i.e. $\mathbf{SO}(\mathbf{3})$ FT (SOFT) which is applicable only for an equiangular grid on a single sphere $\mathbf{S}^2$. Using SPFT, we can use the spectrum information present not only in one spherical layer at a radius but on all radii. Given two rotationally related volumes, i.e. one is rotated version of another, their multi-layered spectra computed from respective SPFTs can be used for correlation of corresponding spherical images with the SOFT algorithm significantly improving the robustness of the rotation estimation, which is precisely the technique that we propose, analyze and demonstrate in this paper. When compared to a similar method [31] (which has only limited range of estimation $\pm 30^0$) we have the ability to robustly estimate arbitrarily large rotation angles on all three axes at the same time. A fine feature obtained from such a combined use of SPFT and SOFT. The use of multi-radii spherical data from concentric grid representation that carries redundant rotational information has been used in data driven analysis in [31,46], as well as learning based approach in [21] that uses spatial occupancy grid based concentric spheres instead.

The rest of this paper is organized as follows: In Section 2 we briefly discuss some related work and the motivation for our approach. Section 3 first shows the robust 3D registration solution for up to 7 DOF transformation using SPFT and SOFT techniques, then we describe the pseudo-code and the computational complexity. We also briefly mention a direct extension of the rotation estimation scheme useful for higher dimensions. Section 4 contains extensive experimentation using volume/voxel data with hundreds of 3D objects used for registrations. We compare the performance of the proposed method with three traditional methods [30,31,34]. And for machine learning using various networks, where some that are designed to exploit spherical properties based techniques, Esteves et al. [18] and Mukhaimar et al. [21], along with others for the robust object classification task. Finally, we follow by a summary of contributions and conclusions in Section 5.

In the spirit of reproducible research, source code and some supplementary materials is made openly available at: https://github.com/syedalamabbas

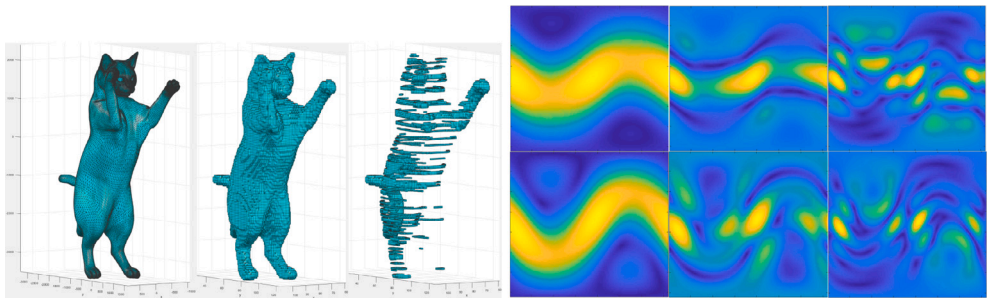

Fig. 1. Left, surface data is transformed into voxels or volumetric representation using suitable resolution then holes are added, overlap between the original cat model and the distorted model is just about 36 percentage i.e. only a small proportion of matching voxels exist even with a perfect alignment (for simplicity no rotation here). Right-Top row, SPFT based magnitude spherical images in increasing order of radii using original cat voxel model. Right-Bottom, SPFT data of the distorted model.

2. Related work

Within the deep learning literature [47], analyzing 3D or even higher dimensional data, can be in general divided based on the three different representations used for the input data, namely the surface representation often used in computer graphics, volumetric/voxels representations [21,48–50], and point cloud representation, Bronstein et al. [51], Qi et al. [52], Poulenard et al. [53], Wang et al. [54], Mukhaimar et al. [55] and Thomas et al. [56]. Alternatively, into two broad categories, those methods that operate on non-Euclidean input such as surface data or point cloud, and those that require Euclidean space input such as regular grid voxels. The most natural extension of 2D Euclidean spaces, methods and the used networks to 3D is to simply use the volumetric or Euclidean representation. For example a voxel-grid representation of the 3D object [57] can be used to amend the 2D CNN framework, and then use the collections of 3D filters for cascaded processing in place of conventional 2D filters. Often the non-Euclidean inputs such as point cloud or surface data have been approximated by Euclidean representations or voxel/volumetric spaces [49], and in some applications data that exist are natively volumetric e.g., volume imaging of magnetic resonance data (MRI), that collects several contiguous slices through a region of imaged object volume rather than a single tomographic slice. However, there has also been a continued interest and parallel development of methods that operate directly on non-Euclidean domain such as structural information being extracted using graph-based methods [58] or learned for manifolds or graphs such as graph convolutional networks [51]. There is a third approach that uses 2D image data for 3D shape analysis. For a clean dataset, the most successful approach is view-based [59], operating in rendered views of the 3D object. The high performance of this method is attributed to the use of large pre-trained 2D CNNs that successfully operate on planar images e.g., ImageNet. However, as we show in our study of classification, in Section 4.7 with ModelNet40 data [57], a number of these networks suffer significantly in the presence of noise and other forms of data corruptions, whereas using our approach with SPFT we are able to achieve a robust classification performance.

We briefly discuss the motivation for using SPFT and some of its properties. This operation first transforms the Euclidean 3D spatial voxels to spherical polar grid in the spectral domain. The spherical spectral signatures on concentric spheres thus obtained have certain special properties such as smoothness, symmetry, and structure. They are arranged in the form of low to high frequencies corresponding to increasing order of radii, see Fig. 1. The overall idea here is that perturbations and missing data, even when they are severe in the spatial domain, have their effect globally diffused and diluted in the spectral domain due to the inherent properties of such a transformation. Hence even with significant distortion of the signal spatially, the redundant spectral signatures at multiple radii have sufficient similar features retained for operations such as a global spherical correlation or convolution to continue to remain effective. During this transformation, not only the domain of a signal is changed but also its coordinate system is modified. The reasoning then is, if SPFT (first stage FT in the Euclidean 3D space) gives us robustness to noise and other perturbations, then tools such as spherical correlations/convolutions and SH (second stage FT on multiple spheres) grants us ways to counter the rotational nuisance of such signals in classification, recognition, alignment or other such tasks. This built-in immunity to deviations as well as redundancy to estimate or neutralize rotations (through possible equivariance) makes for it to be a potentially powerful tool.

3. The proposed method of robust 3D registration using SPFT and SOFT

Notation: Lowercase/uppercase letters are used for spatial/frequency domain functions, respectively.

For completeness, we remind the reader of the Fourier **n**-D affine theorem, Bracewell and Bracewell [22] and Papoulis [60], as the special case of similarity investigated in this paper is derived by simply limiting the affine matrix to 7 DOF. An affine transformation is a non-singular linear transformation followed by a translation [61] in 3D using homogeneous coordinates it has the form,

$$X' = \begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix} X \tag{1}$$

where $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \in \mathbf{R}^3$, and $t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$.

Let $f_1(X)$ and $f_2(X), X \in \mathbf{R}^3$, be scalar signals, associated with two volumes i.e. the luminance/intensity functions, which relates through an *affine transformation* $H_A = \{A, t\}$,

$$f_2(X) = f_1(AX + t) \tag{2}$$

Now let us consider the corresponding relationship between the FTs of the two volumes, $f_1(X)$ and $f_2(X)$, using the **n**-D Affine theorem in the above equation we get,

$$F_2(\xi) = \det(A^{-T}) F_1(A^{-T}\xi) e^{-i2\pi\langle \xi, t\rangle} \tag{3}$$

where, ξ is the Fourier domain Cartesian coordinates.

An advantage offered by the consideration of affine transformation, H_A in Fourier domain is that the estimation of matrix A can be decoupled from the estimation of translational vector t. The separability of affine matrix A is achieved by using the magnitude equation,

$$|F_2(\xi)| = |\det(A^{-T})| |F_1(A^{-T}\xi)| \tag{4}$$

$\det(A) > 0$.

3.1. Class I: Isometries

Isometries are transformation of the plane in $\mathbf{R}^3$ that preserve Euclidean distances [61]. When the affine parameters reduces to just

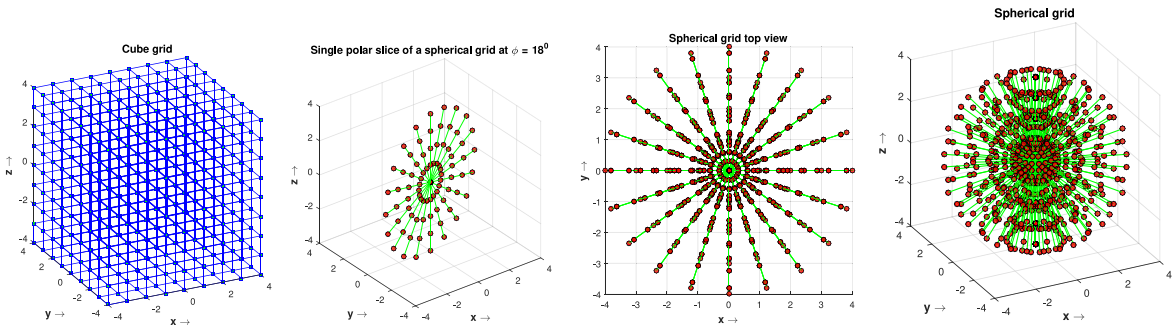

Fig. 2. Given a 3D volume/matrix $f(r,c,d)$ we require transformation from spatial domain Cartesian coordinates of size $(N+1)^3$, N being even, to the Fourier domain spherical coordinates grid $F(k,m,n)$, of size $K \times M \times (N+1)$ where the origin is included multiple times. In the above demonstration $N = 8$ i.e. this is volume of size 9^3, the spherical grid is $10 \times 10 \times 9$ points. $K = M = N+2$, hence azimuth/zenith angular spacing $= 180/K = 18°$.

a rotation $g \in \mathbf{SO(3)}$, it is an isometry (except reflections) represented by,

$$g = A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{5}$$

In $\mathbf{R}^3$, any rotation can be expressed as the composition of three rotations about the three axes x, y, and z. We may also express any element of the rotation group $g \in \mathbf{SO(3)}$ in terms of the rotations in only about the z and y axes using the same conventions as used in [45]. Let,

$$R_z(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix},$$
$$R_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \tag{6}$$

where $0 \le \alpha < 2\pi$ and $0 \le \beta \le \pi$. Using these two matrices, any rotation $g \in \mathbf{SO(3)}$ has an associated *Euler angle decomposition*:

$$g = g(\alpha,\beta,\gamma) = R_z(\alpha)R_y(\beta)R_z(\gamma) \tag{7}$$

Thus, $\det(A) = \det(g) = 1$ in (3). This decomposition provides a natural coordinate system for working with functions on $\mathbf{SO(3)}$, where a function $\Lambda(g)$ for $g \in \mathbf{SO(3)}$ is given by $\Lambda(\alpha,\beta,\gamma)$ where $0 \le \alpha,\gamma < 2\pi$ and $0 \le \beta \le \pi$ (see Fig. 2).

Given the Cartesian 3D equispaced volume data $f(r,c,d)$, where r,c,d represent rows, columns and depth indices respectively, we define the *Spherical Polar* or *Spherical* FT (SPFT) to be the collection of samples $F(k,m,n)$, given by,

$$F(k,m,n) = \sum_{d=-N/2}^{N/2} \sum_{c=-N/2}^{N/2} \sum_{r=-N/2}^{N/2} f(r,c,d)$$
$$e^{-i\frac{2\pi n}{N+1}(r\cos\theta_k\cos\phi_m + c\cos\theta_k\sin\phi_m + d\sin\theta_k)} \tag{8}$$

where, $-N/2 \le n \le N/2$ represents the points on a radial line, $\theta_k = k\Delta\theta, 0 \le k \le K-1$ represents the *zenith* angles, $\Delta\theta = 180°/K$ is the angular spacing in zenith angles, $\phi_m = m\Delta\phi, 0 \le m \le M-1$ represents the *azimuth* angles, $\Delta\phi = 180°/M$ is the angular spacing in azimuth angles, and $K \times M$ is the total number of radial lines. Thus the total number of DFT points is $K \times M \times (N+1)$, where the origin is included multiple times. A fast and exact algorithm with complexity,

$$O\left(3 \times \frac{M}{4} \times \frac{K}{4} \times \frac{(N+1)^3}{2} \log_2(N+1)\right) \tag{9}$$

to compute (8) is proposed in [44], comparatively a brute force approach has a prohibitive complexity of $O((N+1)^6)$, which is essential for the rotation estimation in 3D volume registration discussed below using spherical coordinates.

For the two 3D volumes, f_1 and f_2, related only by a 3D rotation, we can compute the respective discrete SPFTs using (8). This gives us two sets of spherical grids in the Fourier domain, F_1 and F_2, that are also related by the same 3D rotation. Consider just the spectral spherical magnitude data,

$$H_n^1(k,m) = |F_1(k,m,n)|$$
$$H_n^2(k,m) = |F_2(k,m,n)| \tag{10}$$

where $1 \le n \le N/2$ represents the spectral radii, and k,m represent the zenith and azimuth angle indices respectively. Let $\mathbf{L}^2(\mathbf{S}^2)$ denote the space of square integrable functions on sphere $\mathbf{S}^2$. The general idea after the above computations is to use these two spherical images $\{H_n^1(k,m), H_n^2(k,m)\} \in \mathbf{L}^2(\mathbf{S}^2)$ obtained from the magnitude 3D spectrum for correlation [45] in order to determine the 3D rotation angles. The two spherical images that are taken from different radii n from (10) are considered mapped on to a unit spheres. Fig. 3 depicts the proposed multi-layered spectral approach to the rotation estimation problem. Also observe that for the determination of the rotation we only use the 3D magnitude spectrum (10) which is fully decoupled from the overall translation separately computed in Section 3.3.

Given the two functions on the spheres, H_n^1 and H_n^2, at a specific radius, i.e. where n is a constant integer, and the knowledge that H_n^2 is a rotated version of H_n^1, i.e. $H_n^2 = \Lambda(g)H_n^1$ for some $g \in \mathbf{SO(3)}$, the goal is to find g. Here for each $g \in \mathbf{SO(3)}$, we can associate a linear operator $\Lambda(g)$ which acts on $H_n^1 \in \mathbf{L}^2(\mathbf{S}^2)$. Since $H_n^1, H_n^2 \in \mathbf{L}^2(\mathbf{S}^2)$ are two band limited functions on the spheres, we can compute the Fourier expansions of these functions in terms of the SH functions Y,

$$H_n^2(\omega) = \sum_p \sum_{|q|\le p} \mathbf{A}_{pq} Y_p^q(\omega)$$
$$H_n^1(\omega) = \sum_p \sum_{|q|\le p} \mathbf{B}_{pq} Y_p^q(\omega) \tag{11}$$

The terms $\mathbf{A}_{pq}$ and $\mathbf{B}_{pq}$ represents the SH coefficients from the corresponding spherical data H_n^2 and H_n^1 respectively. Evaluating g or determining the maximum g (for correlation) by means of FFT on $\mathbf{SO(3)}$ can be achieved by correlating the two spherical functions as follows,

$$\mathbf{C}(g) = \int_{S^2} H_n^2(\omega)\overline{\Lambda(g)H_n^1(\omega)}d\omega \tag{12}$$

where $\overline{\Lambda(g)H_n^1(\omega)}$ denotes the complex conjugate of the term $\Lambda(g)H_n^1(\omega)$.

Using the functions H_n^1 and H_n^2 in terms of SH computed from (11) in (12) and expressing the spherical harmonic of degree p in terms of *Wigner-D function* (used to express the linear combinations of the SH of the same degree), simplifying gives the result,

$$\mathbf{C}(g) = \sum_p \sum_{|q|\le p} \sum_{|q'|\le p} \mathbf{A}_{p-q}\overline{\mathbf{B}_{p-q'}}(-1)^{q-q'}\mathbf{D}_{qq'}^p(g) \tag{13}$$

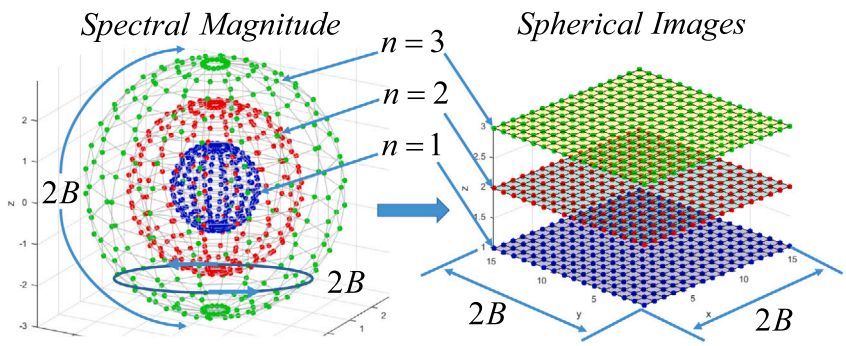

Fig. 3. Multi-layered approach to the 3D rotation angles estimation. Unwrapping of the spherical magnitude equiangular spectrum data computed from SPFT into $2B\times 2B$ equispaced spherical images to be used for correlation using SOFT. In the above demonstration, $B = 8$, $K = 2B$, and $M = B$.

The $\mathbf{D}^{p}_{qq'}$ is a Wigner-D function based on Jacobi polynomials. The correlation can then be efficiently evaluated in terms of Wigner-D functions and using the inverse **SO(3)** FT. A fast and exact algorithm using FFT can be used for correlating two equiangular spherical images, the details of computing (13) using SH is given in [45]. The original data mapped on the sphere is given by the bandwidth B which defines the angular spacing of the points. The search space **SO(3)** is very large and naive optimization approaches will be time consuming, but by choosing a square planar image of size $2B\times 2B$ interpreted as data living on the equiangular $2B \times 2B$ grid on the sphere as shown in Fig. 3, we can efficiently evaluate the correlation $\mathbf{C}(g)$ of H^1_n and H^2_n, at a whole range of g's based on the discretized grid through the use of SH.

Using the SPFT of the two volumes, F_1 and F_2, and by choosing $K = 2B$ and $M = B$ in the computation of (8) we get an equiangular gridded spectral data on the spheres, H^1_n and H^2_n, of size $2B \times 2B$ at different radii n to do the correlation of spherical images using (12). The result of the correlation is the estimate for the angles α_n, β_n and γ_n at a specific spectral radius n. An estimate from a single radius is usually sufficient in the case of full overlapping volumes as shown in our simulations, see Sections 4.1 and 4.4. However, it is known that the spectral signature at some unknown radius will become unreliable due to distortions such as missing data, noise, and partial overlap between the two volumes, see [31]. Since the spectral magnitude data is available at all radii, i.e. $1 \le n \le N/2$, we can compute the estimates of rotation α, β and γ that are supposed to be same for all radii. This built-in redundancy gives us the robustness in the parameter estimation of the rotation matrix using the two volumes as confirmed by the simulations, see Section 4.4.

The original spherical correlation based rotation estimation [45], and the subsequent development using SH [30] for 3D global rotation solution, works only using a single layer, i.e. the points in the volume are mapped onto a sphere, and then the SOFT/SH procedure is computed for the whole 3D object just once. In 3D and for the higher dimensions in general, much of the shape information is spread across multiple layers and integrating/accumulating or remapping the whole volume only onto a single sphere destroys much of the fine structural contents, which in turn leads to poor and erroneous rotational estimates especially in the presence of high noise. Ref. Bülow and Birk [31] has a multi-layer spectral approach similar to the fundamental principle of our method, however there are the following major differences:

- It involves spectral resampling i.e. mapping of 3D Cartesian Fourier magnitude data only onto a set of concentric hemispheres, which is based on interpolations in the Frequency domain. The resampling/projection leads inevitably to deviations from the true rotation. In the proposed method, we avoid this and compute the full 3D SPFT directly from the given volume using a fast and exact algorithm that we presented in [44], hence we are able to capture very accurately the true 3D rotation.
- It requires first the estimation of yaw angle separately with resampling of Fourier data from Cartesian to spherical coordinates, the evaluation of which is crucial for the subsequent full 3D rotation angles computation to work. Our technique is based on the SOFT method of correlating spherical images and hence we get all three angles in one step.
- It uses for roll/pitch rotations a perpendicular projection of the surface of the hemispheres to transform the problem into a 2D shift in matrix form of the projected structures. In our method, its true 3D.
- Due to the above roll/pitch projection based solution, an assumption which holds true only for a restricted range, it is constrained to estimate only small angles i.e. it works as long as roll/pitch both change by at most ±35° each at the same time, where as for our technique the arbitrarily large 3D rotation angles $\alpha_n, \beta_n, \gamma_n$ at a specific radius n can be estimated without any loss of accuracy. The accuracy only depends on the chosen bandwidth or equiangular grid resolution B which in turn also defines the angular resolution of SPFT.

However, due to all such simplifying assumptions and approximations, such as the restricted range, use of only 3D FFT (Euclidean), and the use of phase matching, the SRMR technique in [31] is comparatively faster than our method, proportional to our computational complexity as discussed in Section 3.4.2.

3.2. Class II: Similarity transformation

Let $s > 0$ be the scalar representing the isotropic scaling between the two volumes, then the determinant in (4) becomes, $\det(A) = s^3$, and the scaling parameter can be directly obtained from the ratio of the dc/average values of the two volumes i.e.

$$s \approx \sqrt[3]{\frac{F_2(0,0,0)}{F_1(0,0,0)}} \tag{14}$$

With the new scale compensated volumes, the similarity reduces to an isometry, where the solution can be obtained as described in Section 3.1.

3.3. Estimating the translational vector, t

Consider two volumes, f_1 and f_2, that differ only by a translational vector $t = (t_x, t_y, t_z)$ i.e.

$$f_2(x, y, z) = f_1(x - t_x, y - t_y, z - t_z) \tag{15}$$

For the spatial domain relation in (15), the corresponding FTs, according to the *Fourier shift* theorem are as follows,

$$F_2(u, v, w) = e^{-j2\pi(ut_x+vt_y+wt_z)} \times F_1(u, v, w) \tag{16}$$

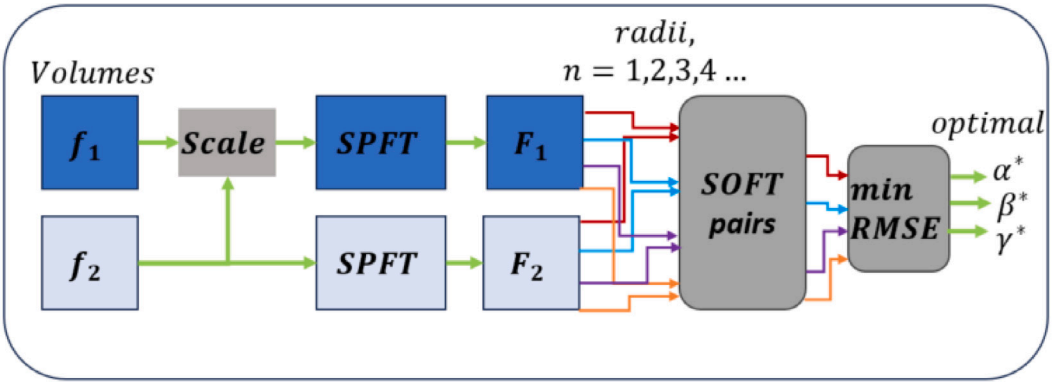

Fig. 4. The basic pipeline of the proposed method. Estimations of all three rotation angles simultaneously occurs not just for a single radius but at various radii before the optimal selection is conducted based on a criteria.

The normalized cross power spectrum is then given by,

$$Q(u,v,w) = \frac{F_2(u,v,w)\overline{F_1(u,v,w)}}{|F_2(u,v,w)\overline{F_1(u,v,w)}|} = e^{-j2\pi(ut_x+vt_y+wt_z)} \tag{17}$$

where $\overline{F_1(u,v,w)}$ stands for complex conjugate. It can also be seen from (3) that the translation information is only encoded in the phase even when the volumes are rotated and scaled. Taking the inverse Fourier transform of (17) we get,

$$q(x,y,z) = \mathbf{F}^{-1}\{Q(u,v,w)\} \tag{18}$$

Then the translation can simply be computed by,

$$(t_x,t_y,t_z) = \underset{x,y,z}{\operatorname{argmax}}\ q(x,y,z) \tag{19}$$

The above solution (19) is called the phase correlation method or phase only matched filtering (POMF) [23,25,60,62], and it is directly applicable to the formulation in higher dimensions. There are a variety of methods to improve this phase correlation to sub-pixel/sub-voxel accuracy, such as Foroosh et al. [25], Shekarforoush et al. [23], Balci and Foroosh [26], Guizar-Sicairos et al. [27] and Tzimiropoulos et al. [28].

3.4. Robust 3D registration algorithm

Fig. 4 shows the block diagram of the proposed method where multi-radii spherical spectrum is used to counter and mitigate the effects of data corruptions. We use a simple metric for the quality of estimated rotation angles, the root mean squared error (RMSE) of the two volumes after alignment. We have chosen RMSE due to its simplicity for demonstration purposes, however in practice a more robust measure like L1/Huber would be used. This measure is used to select the best rotational parameters from the redundant set of estimates of α,β,γ computed at various spectral radii. The RMSE for the two volumes, f_1 and f_2 is defined using the $\mathbf{L_2}$ norm as follows,

$$\text{RMSE} = \frac{1}{(N+1)^3}\|gf_1 - f_2\|_2^2 \tag{20}$$

3.4.1. Pseudo-code

The following is the full pseudo code for our proposed 3D robust transformation estimation of 7 DOF:

Input: Two volumes, f_1, f_2, of size $(N+1)^3$, where N is even. For SOFT procedure, the equiangular bandwidth B, where B is even, and the degree D of SH to be used internally in the SOFT program. For SPFT algorithm, the angular resolution depends on B i.e. $K = 2B$, $M = B$ for the zenith and azimuth angles respectively, where K, M are also even.

Output: Global scale factor, s. Rotational vector, (α,β,γ). Finally, the translational vector, (t_x,t_y,t_z).

Step 1: Compute the average dc values to obtain $F_1(0,0,0)$ and $F_2(0,0,0)$. Find scale factor s according to (14).

Step 2: Compensate for scale s in f_1.

Step 3: Compute SPFT for the two new scale compensated volume pairs, f_1 and f_2, to obtain F_1 and F_2, from definition (8).

Step 4: For the spectral radius n, where $1 \le n \le N/2$, do the following.

Step 5: Gather the two spectral magnitude data, H_n^1 and H_n^2 as defined in (10).

Step 6: Compute the SOFT based correlation of the spherical images H_n^1 and H_n^2 to evaluate the corresponding, $g_n(\alpha_n,\beta_n,\gamma_n)$.

Step 7: Compute the RMSE from (20).

Step 8: Gather the computed estimate at the spectral radius n of Euler angles $\alpha_n,\beta_n,\gamma_n$ and RMSE_n in a matrix of size $(N/2)\times 4$.

Step 9: If $n > N/2+1$, terminate. Else $n = n+1$, go to Step 6.

Step 10: Select minimum error RMSE_n from the matrix and the corresponding angles at n as the final rotational vector.

$$(\alpha^*,\beta^*,\gamma^*) = \underset{\text{RMSE}_n}{\operatorname{argmin}}\ (\alpha_n,\beta_n,\gamma_n,\text{RMSE}_n)$$

Step 11: Compensate for rotation g in f_1.

Step 12: Compute 3D FFTs, $\hat{F}_1$ and $\hat{F}_2$.

Step 13: Calculate the final translation vector, (t_x,t_y,t_z) using (19).

3.4.2. Computational complexity

One of the requirements in the procedure described is the use of SPFT algorithm, where according to (9) (details discussed in [44]), the complexity is,

$$O\left(3\times\frac{M}{4}\times\frac{K}{4}\times\frac{(N+1)^3}{2}\log_2(N+1)\right) \tag{21}$$

This is required for the two volumes. The efficient algorithm implementation for the numerical computation of FTs of the band-limited functions defined on a rotation group $SO(3)$ [45], is presented in [63]. For the pair of spherical images H_n^1 and H_n^2 with bandwidth B, the computational complexity is $O(B^4)$, where the SH is of degree at most B. There are $N/2$ such spectral spherical magnitude data pairs, we thus have the total complexity of using the SOFT technique as,

$$O\left(2\times\frac{N}{2}\times B^4\right) \tag{22}$$

There are two 3D FFTs also involved with the complexity $2 \times (N+1)^3 \log_2(N+1)$ Since we choose $K = 2B$ and $M = B$, the computational complexity of SPFT algorithm dominates the entire 3D robust estimation procedure. Hence, the complexity of the entire registration algorithm is,

$$O\left(6 \times \frac{B}{4} \times \frac{B}{4} \times \frac{(N+1)^3}{2} \log_2(N+1)\right) \tag{23}$$

Due to sufficient computational complexity the proposed solution is more suitable for an off-line rotational estimation application such as sensitive medical data, or a challenging dataset where there is significant amount of distortion such as holes, partial overlaps, noise, etc., where local techniques like ICP are not easily applicable or reliable. We demonstrate the robust nature of the proposed solution in our simulations, see Section 4.5. However, as it can be seen the proposed method is globally robust, non-iterative (fixed complexity) and the fully parallel algorithms of SPFT and SOFT can take full advantage of commodity GPGPU processing hardware. Most importantly each of the components used in our solution has a parallelizable structure and scales naturally with processing and memory resources [44,45]. Also, to reduce the above complexity, we can consider only a few of the spectral spherical layers for the SPFT and SOFT computations rather than many concentric spheres of spectral information. In Section 4, for all experiments, we have used only up to $n = 11$ layers for robustly estimating the rotation angles, regardless of the volumetric resolution.

3.4.3. Higher dimensions

The solution proposed for estimation of similarity transformation can be directly applied to higher dimensions $\mathbf{R}^{\mathbf{n}}$, i.e when $\mathbf{n} > 3$. For estimation of similarity of order $\mathbf{n} > 3$, we can use the transformation to $\mathbf{n}$-dimensional *hyper-spherical* coordinate system using the relationship between the two *tensors* As shown in Section 3.1 the high dimensional rotational parameters can be computed using a fast and exact solution of *hyper-spherical* transform (similar to the one proposed in [44]) and the multilayer redundant correlations (similar to the one proposed in [45] for higher dimensions) on the *hyper-sphere*. The translation can then be similarly estimated using phase correlation.

4. Experimental validation and performance evaluation

For registration comparison, we focus only on the following three traditional, but global, methods,

- Robust Estimation using Spherical Harmonics, Althloothi et al. [30].
- Spectral Registration with Multilayer Resampling, Bülow and Birk [31].
- Go-ICP, Global Optimal Iterative Closest Point, Yang et al. [33] and Yang et al. [34].

RESH [30] and SRMR [31] bear some similarity to our proposed solution, being robust methods well designed for global parametric 3D rotation estimation using Fourier domain information. RESH uses SH, and is suitable only for surface 3D data even when it has holes or missing data, while SRMR uses POMF, and can handle volume data with partial overlap, both of these techniques have some limitations when compared to our proposed method, which as discussed in Section 3.1, and demonstrated and confirmed by the simulations here. The Go-ICP method follows a different working principle compared to the original work of ICP [32] which was also local. We compare the results of our method over the Go-ICP in case of a partial overlap and noisy registration data. We have mainly focused on demonstrating the error prone and most difficult part for any 3D transformations which is the 3D rotation estimation because the computations of scaling using (14) and translation vector using (19) are well known to be robust.

For machine learning comparison study, we use a few recent networks, including those that exploit spherical signatures, Maturana and Scherer [48], Qi et al. [52], Riegler et al. [50], Esteves et al. [18], Thomas et al. [56], Mukhaimar et al. [55], Wang et al. [54], Poulenard et al. [53], Thomas et al. [56] and Mukhaimar et al. [21]. The best advantage of our approach is its inherent robustness to various forms of data perturbations/corruptions, and hence we design the experiments for the object classification problem that can clearly benefit from such features and demonstrate its effective use.

Note: All angles are in radians.

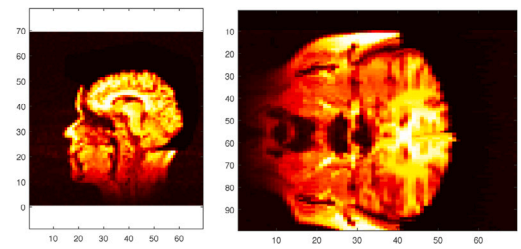

Fig. 5. Slices of the MRI images used to construct a full 3D volume, XY and YZ view of the volume, using the publicly available dataset, Rosenman [64].

4.1. 3D rotation estimation - Simple MRI dataset

We begin with a simple rotation estimation task using the MRI volume with skull removed from the publicly available data set [64]. Figs. 5 and 6 show the reference/fixed volume and the rotated volume, which are visualized through images, slice projections at orthogonal angles and through the surface data. Fig. 7 shows the images obtained using 3D SPFT [44] at radii $n = 1, 2$ and 3, respectively. Notice the slight variations in the spectral signatures for the increasing radii, the sharpness of which will depend upon the volumetric information, especially in the presence of holes as demonstrated later. Using these images we compute the correlations on spheres with the SOFT method [45]. Fig. 8 shows the final 3D rotation estimation results using spherical magnitude data obtained using the two volumes.

The black and white shading on the spheres reflects the wrapped intensities of the spectral spherical images. We also use special marker 3D axes to conveniently visualize the 3D rotational transformation and to visually assist in demonstrating the failed/successful estimation. The axes marked by the stack of green circles is the known actual rotation and the axes marked by the stack of red squares is the estimated one. When these two marker axes (green and red) overlap fully within the angular resolution, it signals that the estimation is successful, any misalignment can also be visualized well using this.

Observe that the estimates at radii $n = 1$ and 2 are close to the actual rotation, while the solution at radius $n = 3$ is a perfect match, up to the angular resolution of the equiangular grid defined by B. While both the techniques [30,31] perform similarly with full and clean overlapping 3D models, our solution involves no approximation or interpolation steps, exploits the redundant Fourier domain information, uses only full 3D correlation, and hence it is more robust to noise/holes/partial overlaps as confirmed by the simulations in the following sections. Also note that its accuracy can be increased to an arbitrary precision by either directly increasing the number of angles defined by the bandwidth B or using sub-pixel enhancement in a small region, cf. [27,28].

4.2. Princeton object database, large set

In order to evaluate the accuracy and robustness of the proposed method for 3D rotation estimation of different 3D objects with varying geometric complexities, we conducted hundreds of experiments using the Princeton Shape Benchmark (PSB) database [65]. PSB is a large database for 3D models collected from different web domains, we used a total of 790 objects for our experimentation. These objects are com-

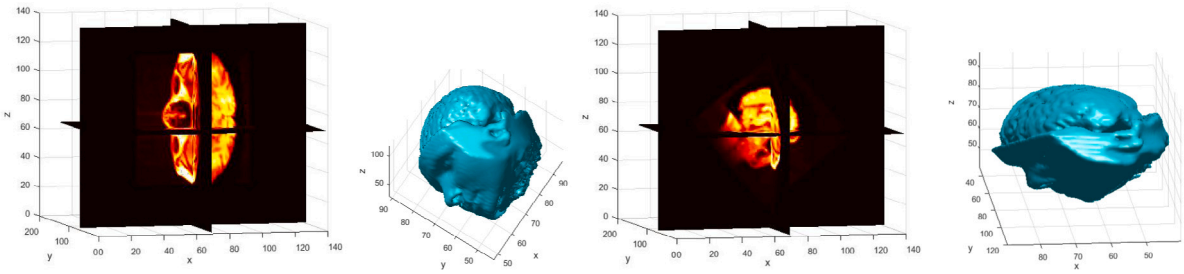

Fig. 6. Slice projections and surface visualizations of the 3D volumes used for comparison with actual rotation matrix in (7) given by Euler angles, $(\alpha, \beta, \gamma) = (0.7069, 1.0996, 1.5080)$. Voxel size for both objects is chosen arbitrarily as $63 \times 63 \times 63$, and the equiangular bandwidth $B = 68$ as discussed in Section 3.1.

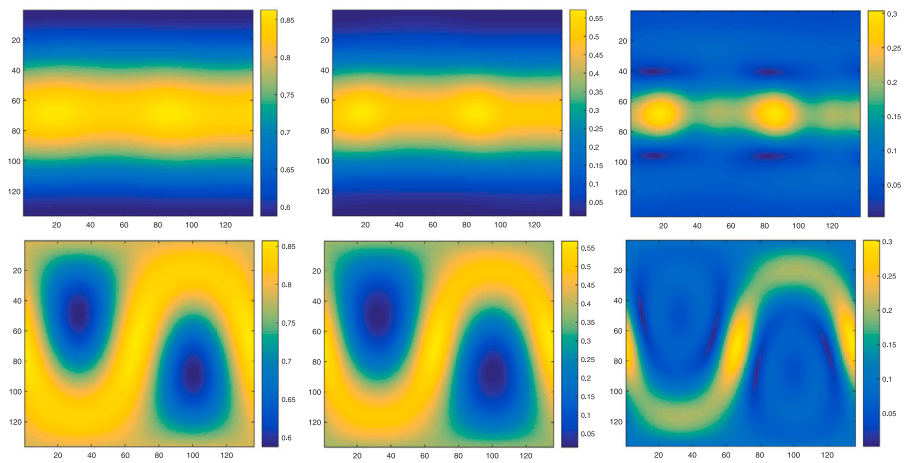

Fig. 7. Spherical images H_n^1 and H_n^2 as described in (10), original (top row) and rotated (bottom row) by $(\alpha, \beta, \gamma) = (0.7069, 1.0996, 1.5080)$, at spectral radii (left to right) $n = 1, 2$ and 3 respectively, obtained from (8) (SPFT) and correlation given by (12) (SOFT). Note the variations in the spectral signature as we go from smaller to larger radius which is equivalent to going from low to high frequencies.

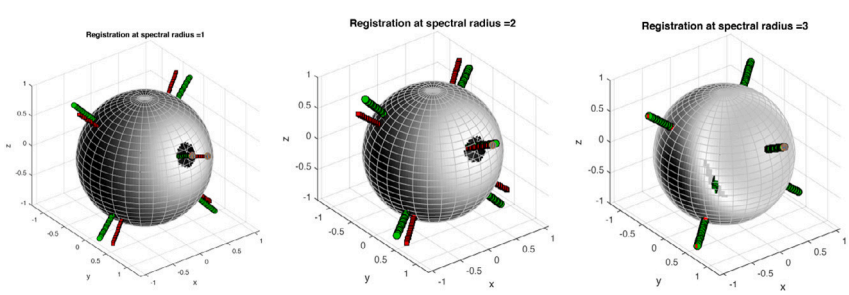

Fig. 8. Evolution of registration at different spectral radii, the estimated parameters at $n = 1$ are $(5.6364, 2.0443, 4.8510)$. Although the solution as observed above is close with $\pi/2\pi$ ambiguity, it is not a perfect match within the angular resolution. The estimated parameters at $n = 2$ are $(2.4024, 1.1972, 1.7632)$ are closer to the ideal one. Finally, at radius $n = 3$ the estimated values $(2.4024, 1.0972, 1.6632)$ are near perfect registration within the angular resolution defined by bandwidth B with π ambiguity for angle α i.e $\pi - \alpha = 0.7392$ which is $\approxeq$ to the actual $\alpha = 0.7069$ radians. Ambiguities are discussed further in the next Section 4.2.

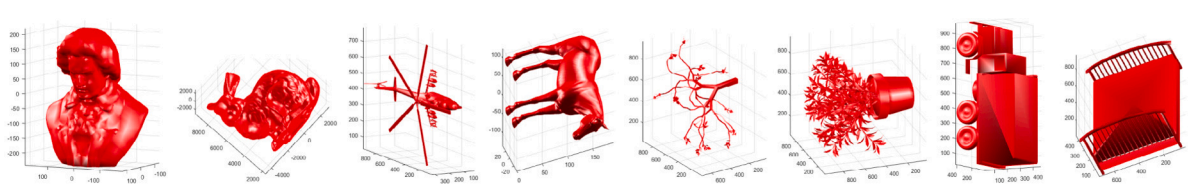

Fig. 9. Different objects with shapes and geometric complexities used in the PSB database test [65], (From left to right) (a) Beethoven (b) Bunny (c) Helicopter (d) Horse (e) Branches (e) Pot Plant (f) Truck (g) Bed.

posed of vertices and faces, which are then voxelized or transformed into volumes using the 3D software package, SPHARM-MAT, Li [66], for the use of our method, see Fig. 9 for few sample objects in their original form.

We compute the RMSE value between the volumes of the rotated object and reference/fixed object in the spatial domain after the rotation estimation, using (20) which gives us a crude global measure for the quality of estimated rotation angles in the case of fully overlapping

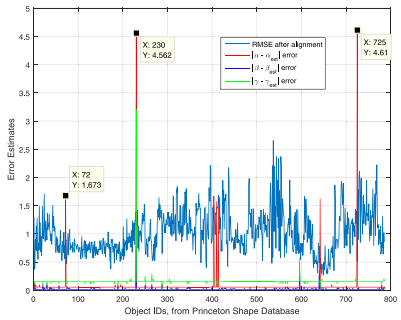

Fig. 10. The error estimates obtained for the PSB objects that are 790 in total, the known fixed 3D Euler angles are $(\alpha, \beta, \gamma) = (2.2777, 1.9018, 1.5080)$ and are used for rotation according to (7). The volumetric resolution for all the objects after voxelization is $63 \times 63 \times 63$.

Table 1
Effects of perturbation on the surface vertices.

SNR (dB)	$\lvert\alpha\rvert_{\text{error}}$	$\lvert\beta\rvert_{\text{error}}$	$\lvert\gamma\rvert_{\text{error}}$	RMSE
10	0.1861	0.0893	0.2547	0.1893
15	0.0858	0.0558	0.1210	0.0915
20	0.0144	0.0057	0.0542	0.0325
30	0.0479	0.0057	0.0207	0.0303
35	0.0190	0.0057	0.0207	0.0165
40	0.0184	0.0057	0.0542	0.0321

volumes. We also compute the absolute error for each of the Euler angles α, β, γ. Fig. 10 shows the plot for all the objects used for this experiment and the error estimates.

The special objects where the angular errors are large, even though the RMSE are not as abnormal, are investigated further. The registration of these few objects is shown explicitly in Fig. 11. The problematic estimations occur for objects that are elongated, largely flat or rotationally symmetric about any axis. However, as it can be seen from Fig. 11, these estimations are largely accurate. The errors are due to the inevitable $\pi/2\pi$ ambiguity of the rotationally symmetric problem, demonstrating that the proposed solution is effective and independent of the shape of the object to be registered.

4.3. Robustness to noisy points

Here we evaluate the robustness of our proposed method. The x, y and z vertices of the original object are perturbed by an additive white Gaussian noise, where SNR varies from 10 to 40 dB. We conduct several experiments using 20 different objects randomly chosen from the PSB database, see Fig. 12 for the demonstration of *Dog* object and Table 1 with RMSE.

We observe that it is very robust to additive noise even when SNR = 10 dB beyond which the object shapes becomes sufficiently distorted to not merit any reasonable comparison. The very crude spectral signatures are largely still sufficient for effectively computing the rotation using it. Thus, it can also be used with objects that have intra-class variations [30]. We also note that RMSE is inherently tied to not just the estimate obtained from our solution but also the level of voxelization and the resulting angular resolution which itself introduces approximation errors.

4.4. Missing voxels or holes or partial data

Here we consider the case when one of the volumes has missing data or holes. RESH requires that the 3D surface (vertices and faces) to be connected regions with no holes. In practice, this topology requirement may not be satisfied due to sensor errors. In order to compute the SH coefficients then an interpolation technique is used. Mesh surface is first converted into voxels, then filled up using a connectivity function, and finally reconverted back to the surface data, see [30] for details. This requires a pre-processing stage of holes filling defined by a user parameter ϵ (holes larger than ϵ will not be filled). It is difficult to choose this parameter, since on the one hand a conservative selection would leave holes, and on the other an aggressive one would render the whole topology altered. Our solution is fully automatic, i.e. no hole filling process and still the performance, as shown, is superior.

Table 2
Performance in case of holes/missing or partial data. The metric used here is angular error as the ground truth is correctly known and due to the presence of holes RMSE alone may not an adequate and reliable measure to showcase the efficacy.

Method	>80%	>60%	>40%	>30%	>20%
SPFT+	0.0980	0.0980	0.1091	0.1030	0.1872
RESH	0.0045	0.6832	1.4060	1.0378	1.4637

We perform our test of holes/missing data on 20 different 3D objects randomly chosen from PSB. Fig. 13 shows the original surface data of the *Beethoven* bust object, which is first voxelized to obtain the noisy transformed surface data. Fig. 14 shows the registration spherical maps for different radii. As it can be seen here, the rotation estimation is close at radii $n = 1, 2, 5, 6$ with $\pi/2\pi$ ambiguity, while it fails for radii $n = 3, 4, 7, 8$ and finally it is very close to the ideal solution for radii $n = 9$ and 10. RESH is only robust to very small localized holes. However, in our case due to significant distortions and the non-localized nature of the holes, it fails. As can be seen in the simulation, due to the missing data, it affects the proposed registration process for different radii differently. The extent of distortion in a spherical layer depends on structural content that is missing in a volume.

Table 2 shows that our method (SPFT+) produces consistently good estimation of the rotation angles even when the overlap (from voxels) (24) reaches up to 30%. Comparatively, RESH is effective only up to 80%, below it the solution is largely random with significant differences in the angular error, and a fine tuning of hole filling will be required depending on the objects and extent of holes to achieve reasonable estimations.

$$\%\text{Overlap} = \frac{\#\text{Original} - \#\text{Missing}}{\#\text{Original}} \times 100 \tag{24}$$

RESH is applicable only for surface data i.e. logical volumes, while SPFT+ is well designed to handle the full volume data, accounting for the intensity variations along with the shape. We use the same volumes and rotation angles used previously in Fig. 6 of Section 4.1, but this time we randomly introduce missing data. Fig. 15 shows the volume data with significant distortions at all layers of the volume. Observe that due to noise, registration is unsuccessful at radius = 3 (previously it was), in fact it fails for radius = $(1, 2, 3, 4, 5)$, but at radius = 6 we get a near perfect estimate, similar to the noiseless one in Fig. 8.

As it can be inferred from the above simulations, the proposed method performs very well even with several broken layers in a volume, making it very powerful. Possible reasons for this robust performance are: the use of exact computations for SPFT, the global nature of FT, the multi-layered approach, and finally the use of full 3D spherical correlations using SOFT.

4.5. Real world challenging dataset: Crashed car parking disaster scenario

A challenging real-world dataset is used here to compare with SRMR. It deals with a disaster response scenario using a robot for inspection of a collapsed car parking-lot in Disaster City, Texas, see Fig. 16. The robot is operated under conditions with high amounts of rubble and dust on slippery surfaces and hence the scans are very noisy. There are a total of 35 scans present, and a total of 35 scan pairs were used to test the proposed method of registration, it is exactly the experimental setup used in [31]. Many of the scan pairs have only small

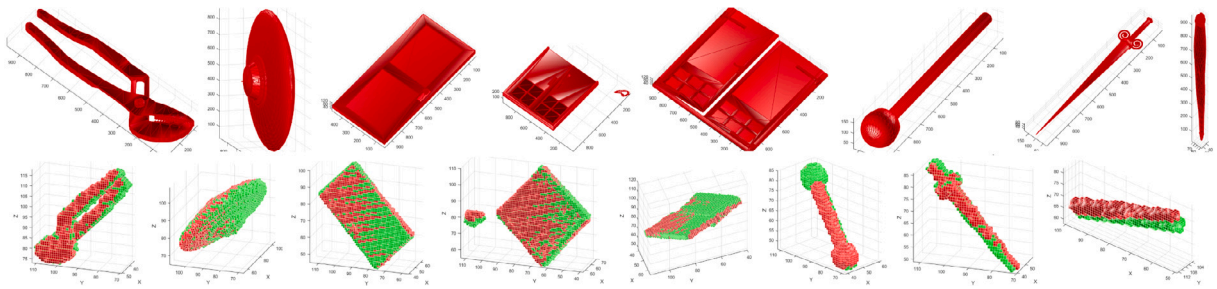

Fig. 11. All the anomalies of the PSB test with the object IDs in $\{72, 230, 404, 410, 414, 596, 642, 725\}$. (Top row) Original objects with large angular errors after registration. (Bottom row) Voxelized fixed and registered 3D objects (red and green respectively) after the registration using our solution. Note the full $\pi/2\pi$ ambiguity for certain objects after registration and the rotationally symmetric shape of other objects. Also notice that only some ambiguities (asymmetric), such as object ID 596 i.e. drumstick flip along its length, can be easily resolved with a suitable spatial metric.

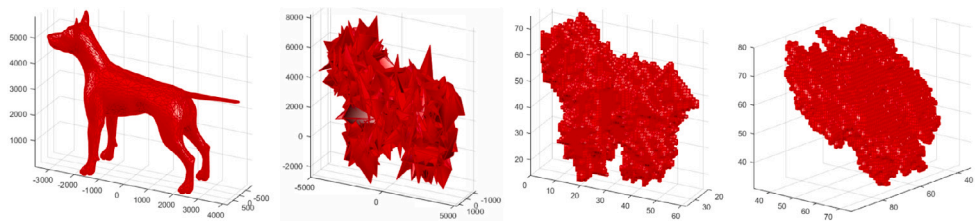

Fig. 12. Surface data transformation for rotation estimation, shown here using *Dog* object from PSB database [65]. Voxel size for both objects is chosen as $89 \times 89 \times 89$, and the equiangular bandwidth $B = 94$. SNR = 10 dB (measured). From left to right: (a) original surface, (b) surface vertices perturbed with 10 dB noise, (c) voxelized and noisy object, (d) voxelized, noisy and rotated object. Actual angles used for rotation are, $(1.0505, 0.7159, 0.7226)$. Estimated angles at spectral radius, $n = 2$ are $(\pi - 1.9050, 0.6266, \pi - 2.6737)$ which comes very close to the ground truth even in the presence of such high levels of distortion.

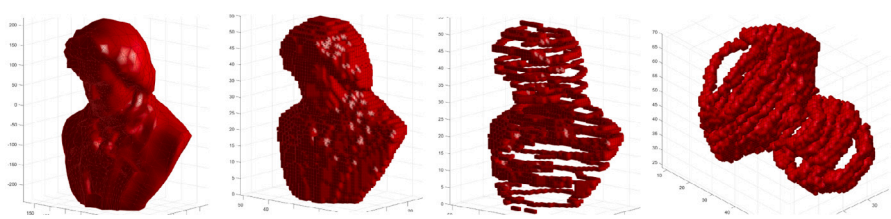

Fig. 13. Surface data transformation for rotation estimation, using *Beethoven* bust from PSB database [65]. Voxel size for both objects is chosen as $75 \times 75 \times 75$, and the equiangular bandwidth $B = 80$. The surface data is first voxelized then significant distortion is introduced at all levels in the form of missing surface data by setting certain values to zeros. From left to right: (a) original surface, (b) pure voxelized object, (c) voxelized and noisy object, (d) voxelized, noisy and rotated object.

overlaps (range 20–70%), and there are many interfering structures, see [31] for details.

The captured data is a set of raw point clouds, hence it requires some preprocessing, i.e. conversion into the volume/voxel form. While it can be argued that this process of conversion from point clouds to voxels would likely lose some accuracy due to lossy transformation, however the structural integrity of the model at a coarse and global scale can be considered intact assuming there is not significant down-sampling. There are different methods of triangulation (from point clouds) and voxelization, but we use the same exact set of volumes as was used in [31] to conduct our experiments. We do not use or process the original form of point clouds here, but for this kind of discrete geometry processing there are software packages with state of the art algorithms available, for instance see [66,67]. Although, we lose some accuracy by this process of transformation of point clouds to voxels due to sub-sampling, nevertheless the overall structure can be well captured by choosing an appropriate voxel size. For our experiments we have used all the scans of equal sizes, $129 \times 129 \times 129$. Once the parameters of 6 DOF transformations (α, β, γ) and (t_x, t_y, t_z) have been computed between the consecutive scans, it can be applied to the original raw form of point clouds obtained directly from the robot sensor, but the overall registration process can be verified just by using the voxel data as we show in our simulation. Since the computed solution is based on global estimates, it gives a good initial (rough) transformation between the consecutive scans. Further refinements can be done using this initial estimate for a more fine local approach such as ICP, which in combination gives more accurate registration results. We also show cases when the local refinements fail despite choosing a good rough initial estimate provided by our solution. The validation of the registration result in this case is done by visual confirmation (see Fig. 17).

A problem with intensity based image/volume registration [68,69] and ICP registration [32] (using MATLAB function *imregister* and *pcrigid*), occurs when there is only a partial overlap between the scans, see Fig. 18. However, when the volumes have sufficient overlap local refinements work well to give good registration, see Fig. 19.

Out of the 6 DOF parameters the rotational vector (α, β, γ) is more difficult to determine, the translation vector solution according to (19) is usually more robust. The full integrated scan of a total of successful 30 scans is shown in Fig. 20. Two of the scan pairs where the registration fails, are shown in Fig. 21. These failed scans are those with largely flat structures with not much vertical (or height of the 3D structure) information, or poor overlap between the two scans. Out of the 35 scan pairs only 3 cannot be registered with the proposed technique i.e. 91.% success, which is competitive with SRMR (94% success rate or 2 failed scan pairs) in this context with a difference of only one. However, as discussed in Section 3.1 there are a number of advantages in using our method compared to SRMR, which is applicable only for small rotation angles. There were many other comparisons reported in [31] using the same exact dataset used, such as ICP-40%, SOFT spectral (single

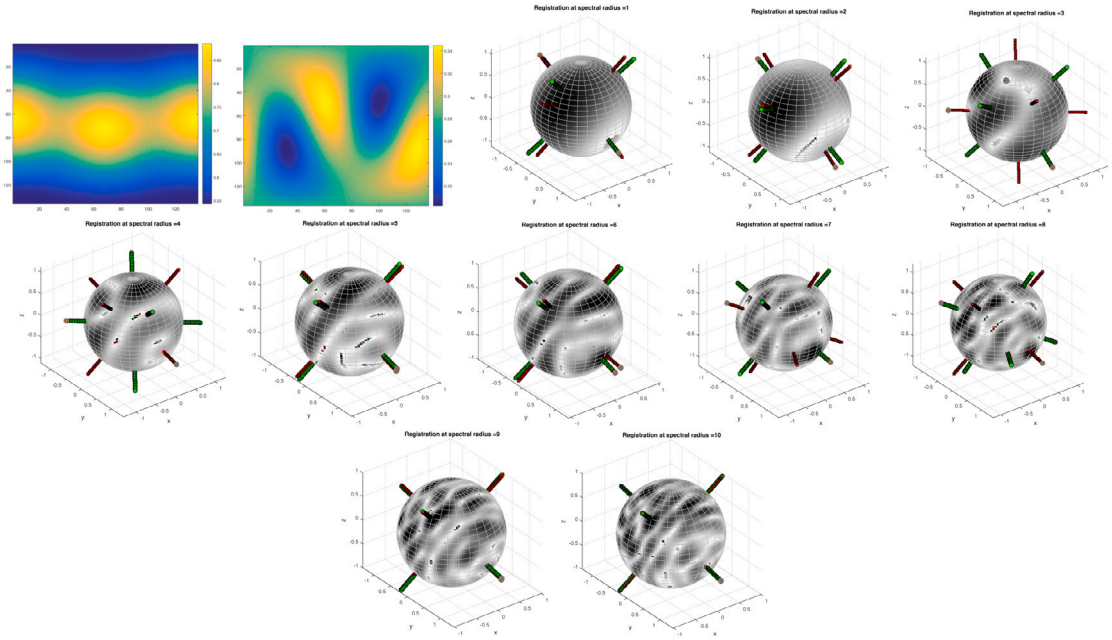

Fig. 14. (From left to right) The spherical images obtained at spectral radius 1 of the original and the noisy transformed surface data. Rotation estimation using the proposed method at various spectral radii from $r = 1 \cdots 10$. The actual solution α, β and γ as discussed in (7) is $= (2.2777, 1.8850, 1.5080)$, and the estimated solution at radius 10 is near perfect match within the angular resolution $= (0.7854, 1.8826, 1.6632)$ again with π ambiguity for angle α.

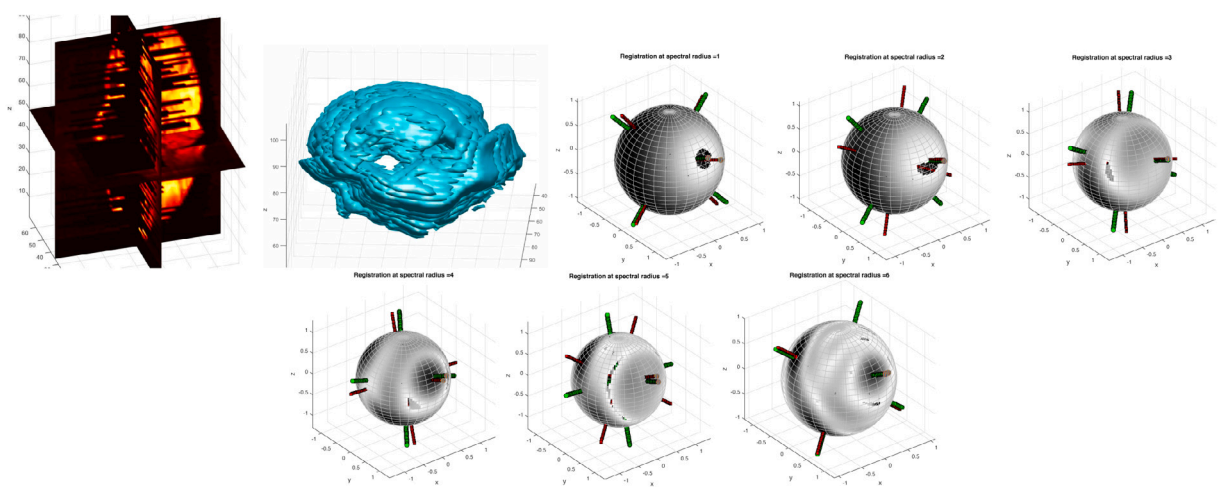

Fig. 15. Rotation estimation with the same 3D volume and the transformation angles same as in Fig. 6. Registration at radius = 6 gives the estimated values $= (2.4024, 1.0972, 1.6632)$, which is near perfect registration with π ambiguity for angle α i.e $\pi - \alpha = 0.7392$, which is approximately to the actual orientation α, demonstrating that the solution is very robust to missing data/ partial overlap due to noise.

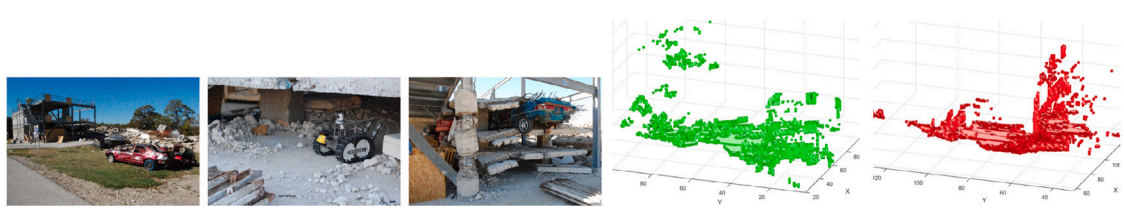

Fig. 16. The Crashed Car Parking Dataset, (left to right) Front view, Close up of robot used and the close up of the site. For instance, the original volume Scan 12 (left in green) and Scan 13 (right in red) displayed in the form of surface data directly constructed from their respective volumes.

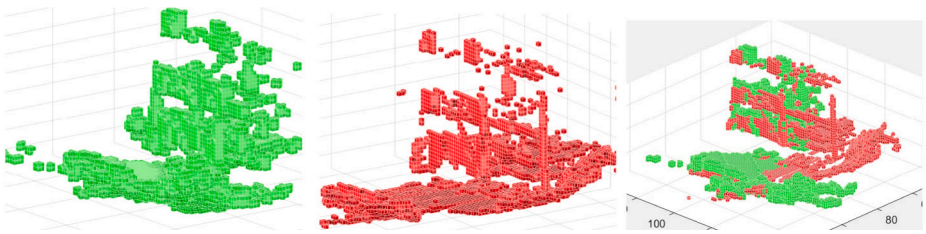

Fig. 17. Volume Scan 12 registered (left in green) and Scan 13 is fixed (middle in red). Right, close up of the combined registered scan directly using the 6 DOF computation using the proposed method. Without any local refinements point-to-point RMSE is a bit higher = 9.456, so it may not be the best metric to use in case of partial overlap.

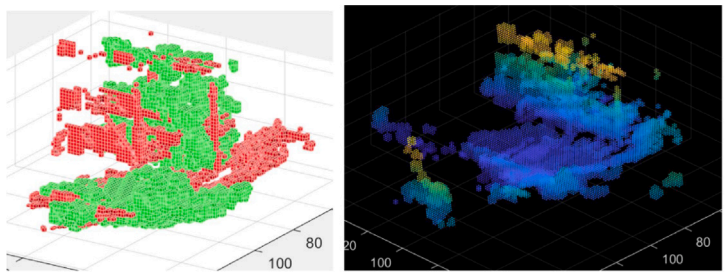

Fig. 18. A direct attempt to register the two volumes using local intensity based (as seen on left) or location based registration ICP (as seen on right) fails. The local refinement registration ICP fails, though RMSE = 8.0968 which is lower, even after the initial estimate is computed using the global proposed method due to partial overlap. However, this demonstrates that roughly the global alignment can be computed using our approach and further direct local refinements are not useful in such scenarios.

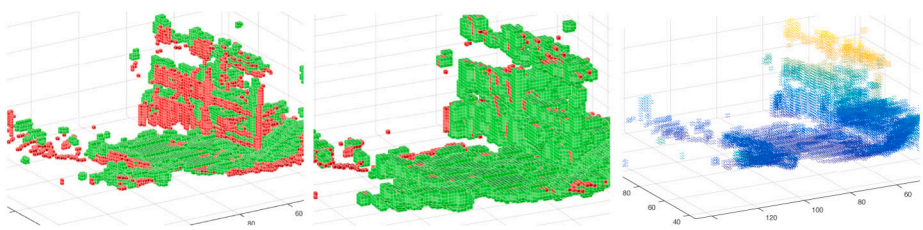

Fig. 19. Successful refinement using local registration is useful for scans with sufficient overlap e.g. Scans 13 and 14, (left) directly registered globally using proposed method, (center) further refined using image intensity based local registration and (right) refined using ICP.

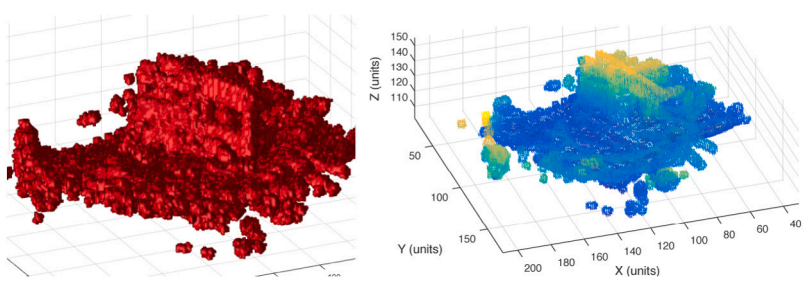

Fig. 20. Final integrated plot of 30 (left and right) the volume data and its corresponding point cloud set, as can be seen all the scans have been well aligned globally (roughly) from the (sub-sampled) volumetric data alone. Further local refinements can be done to improve the alignment by directly using the original point cloud data.

layer) [45]- 83%, plane matching-71%, etc. Interestingly, even a single layer SOFT technique was comparable or second best to the SRMR method, and we have improved (compared to the single layer SOFT in [45]) its success rate and made it more robust by introducing the multi-layered approach in this paper.

4.6. Comparison with Go-ICP method for partial overlap

The global ICP algorithm is a subsequent development on the previously well-known ICP technique [32], which is the most widely used method of 3D registration with point-set data. However, as it is based on local iterative optimization its susceptible to converge on a local minimum. A global optimal solution, Go-ICP, proposed in [34] is based on a branch-and-bound (BnB) scheme that searches the entire 3D motion space $SE(3)$ or 6 parameters.

For comparisons, in order to maintain consistency between the point cloud data used for Go-ICP and the voxel data required for our method, we first voxelize the objects from the PSB database and retain the same vertices that represent the corners of the voxels. Thus the same voxels and the exact vertices of the voxels are used for the two methods for a fair comparison. Fig. 22 shows the registration results for the Bunny object using the Go-ICP, and Fig. 23 shows the same registration using the proposed method, where only 3D rotation transformation exists between the two volumes. For estimating the run time of the Go-ICP method, we repeated the experiment for 20 trials and the average running time estimated = 5377.9 s. Note that the run time here depends

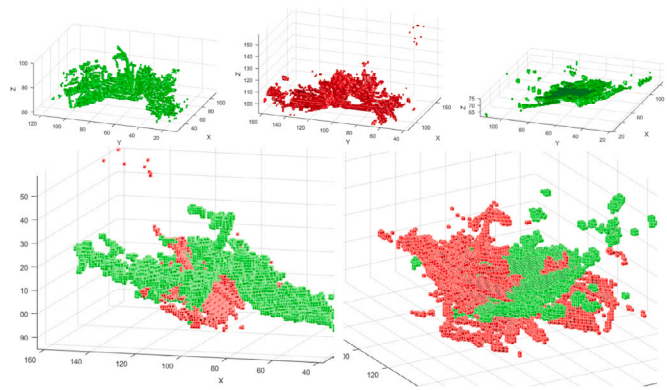

Fig. 21. Failed Scans. (Top row) Scan 1, 2 and 35. (Bottom row, left to right) failed registration for scan pair 1-2 (green–red), failed registration for scan pair 35-1 (green–red).

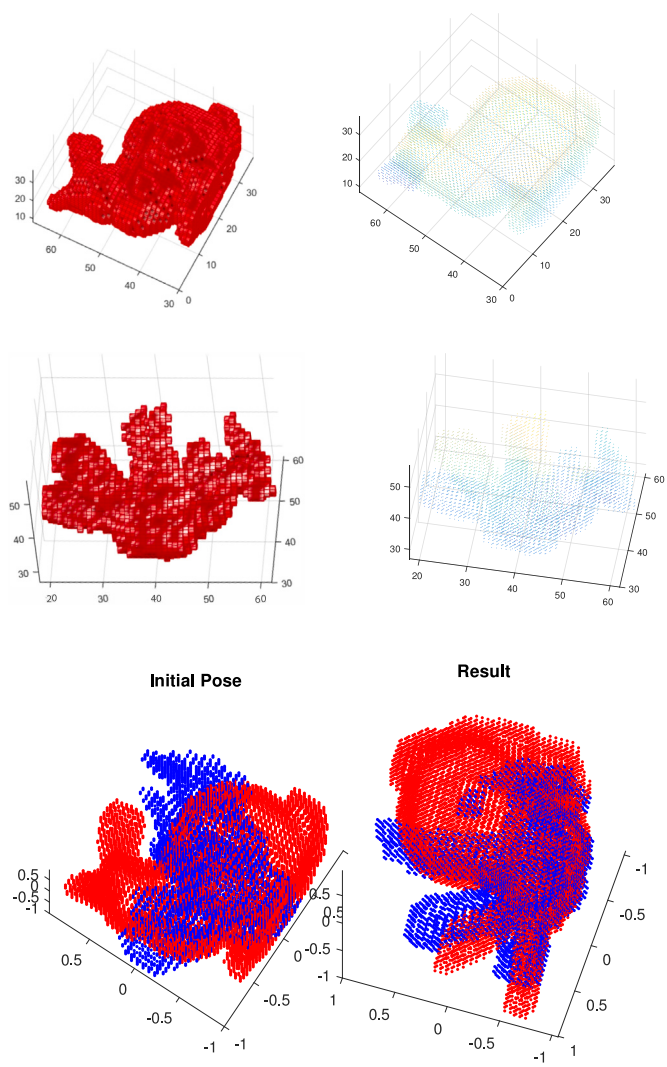

Fig. 22. (Top to bottom, in pairs) (a) Original voxelized bunny object and its point cloud representation (b) Partial and rotated voxelized part of bunny and its corresponding point-set data (c) (viewed from a vantage point) Go-ICP with noisy and partial data, solution. The actual rotation Euler angles are $(1.2794, 1.1718, 0.7226)$ and the Go-ICP estimate $(0.2599, 1.2290, 4.0183)$ are close but not very accurate. The run time here is significant (in hours).

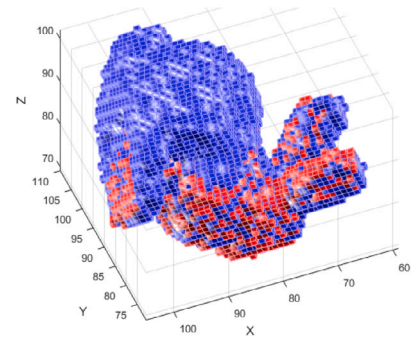

Fig. 23. The angles computed with the proposed solution are very close to the actual one $(1.2566, 1.1192, .7330)$ even when the overlap is just 38.76%. We use $B = 120$, i.e. the two spherical grids (8) are of size $240 \times 120 \times 115$, using an optimized GPU based computation on a commodity NVIDIA GTX 760 hardware.

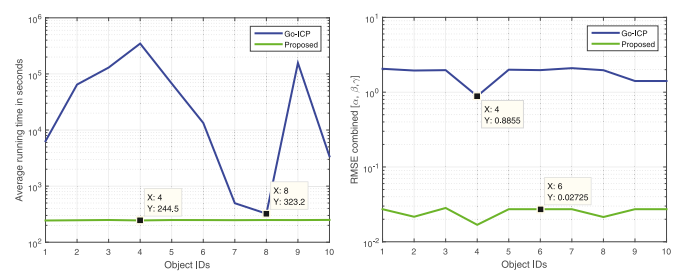

Fig. 24. (Left) Speed comparison (Go-ICP can take an arbitrary amount of time-from hours to days) and, (Right) Accuracy. In the above experiments we fixed the actual rotation vector, $(\alpha, \beta, \gamma) = (1.2794, 1.1718, 0.7226)$, and the two volumes were of size $65 \times 65 \times 65$.

on the signal content and the extent of the overlap between the two objects. There is 38.76% overlap present between the fixed bunny and the rotated partial bunny 3D object. The Go-ICP estimate is not very accurate despite taking a lot of time searching a truly global solution compared to the proposed technique. The run time for the proposed method using the two volumes of size 115^3 are 294.17 and 294.58 s (unoptimized), which is fixed and independent of the object shapes or the overlap between them. For the execution time of the algorithm SPFT and more details, which takes up most of the time in the above procedure, the reader is referred to Abbas et al. [44]. Note that we do not need the entire spherical polar grid for robust 3D registration, only a few spherical layers (spectral radii $n \in [1, 11]$ with image pairs of size 240×120) will suffice, so that in real-world the solution can be computed faster than reported here. For the above experiment, we used only up to 11 spherical layers to choose the most robust estimation of the angles based on RMSE.

A possible reason for the slow processing of Go-ICP maybe due to the exponential increase in the 6D search space with the increase in the number of points and reduction in the overlap. Even if we consider that the current code is not very optimized and hence the execution speed is low, the final estimates are still not as accurate as the proposed solution.

Finally, in case of noise and outliers, the Go-ICP requires careful selection and refinement of statistical outliers before processing, while the proposed solution is obtained without any such refinements but may still benefit from some outliers rejections. Note that there is a parameter selection required by the user for the number of points (vertices) to be used between the two objects which defines the limits of the search space, e.g. from 1000 points to 2000 points the run time increases significantly (2–3 times higher).

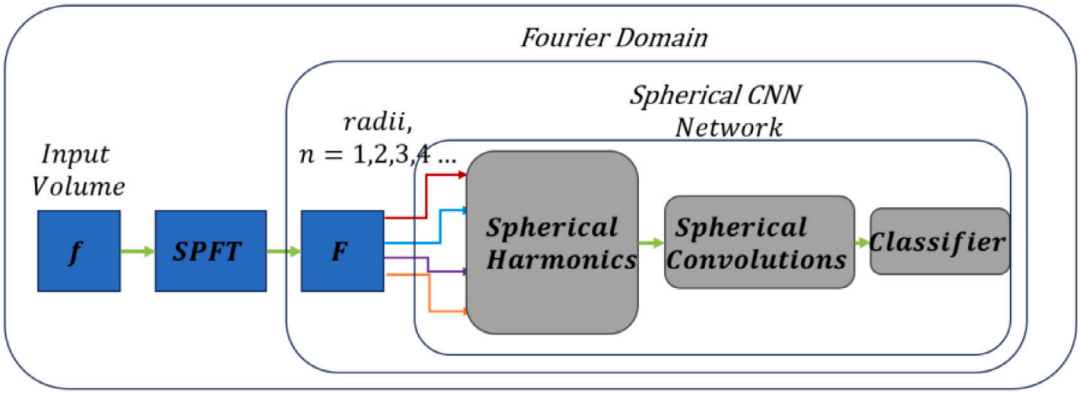

Fig. 25. The basic pipeline of the proposed Spherical CNN architecture. Spectral spherical magnitude data at various radii allows for achieving robust performance in the classification task.

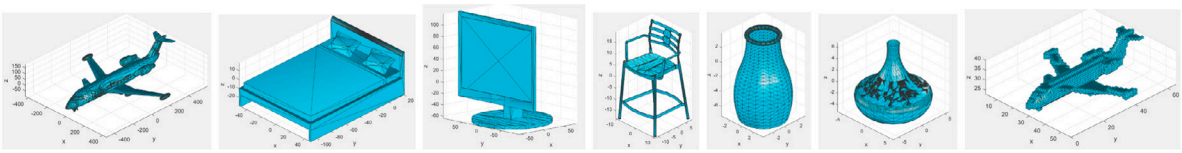

Fig. 26. Six random objects from ModelNet40 dataset, from left to right: airplane, bed, monitor, chair, flower pot and vase. Last rightmost figure shows a sample voxelization of the model airplane which is a necessary prerequisite before the processing of SPFT based spherical CNN architecture can process it.

Thus the proposed method provides a truly global non-iterative solution to the difficult 3D rotation problem, it is faster than the Go-ICP method, has a fixed run time and complexity, and finally is more accurate. We conduct experiments using additional 10 randomly selected different objects from the PSB dataset, the average running time for each experiment (using both methods) with each object was computed in 20 trials. Fig. 24 shows the final comparison results in terms of speed and accuracy, respectively. It can be seen that the Go-ICP method running time can be often unpredictable and its accuracy still is not very competitive. However, it may be more suited to be used directly on point clouds of smaller sizes, arbitrarily sampled data, and lesser distortions.

4.7. Robust object classification, ModelNet40 dataset

In this section, we use spherical images from (magnitude) SPFT to train a spherical convolutional network, and compare it against a few recent classification networks. Our goal is to demonstrate that the multi-radii spherical spectral signatures are robust, reliable and effective for the classification task even in the presence of various types of significant data perturbations/corruptions. Fig. 25 shows a slightly modified, from [21], and adapted spherical CNN architecture diagram for using SPFT spherical images in robust classification task.

We benchmark performance against the ModelNet40 dataset, Wu et al. [57], that has a total of 12 311 objects in surface representation form (triangular simplicial mesh with vertices and faces), see Fig. 26 showing few sample objects from it. The data in it is split into 2468 objects for the test set and 9843 objects are in the training set. We also separate or take about 1000 objects from the training set for the validation set which is necessary during the training sessions. We consider three types of real-world uncertainties in the data [21], missing data in which we randomly drop out points from the voxels that will be used for feeding the network, outliers where we randomly insert points in some proportion, and additive noise where the original vertices are deviated.

We create the following four scenarios, refer to Fig. 27, to perform the classification task:

- Clean training (~8000 objects), validation (~1000 objects) and testing (2468 objects) set.
- Clean training set, and infected training set (again ~8000 objects) with additive noise (20 dB), clean validation set, test set infected with additive noise.
- Clean training set, and infected training set with random dropouts of points (90%) for missing data simulation, clean validation set, test set with missing data.
- Clean training set, and infected training set with un-clustered outliers (50%), clean validation set, test set with infected outliers.

We follow a similar procedure as in previous section to use our method. First we transform the surface representation into a regular voxel grid, see Fig. 27 to visualize them in a regular point cloud form. The volume resolution used is $65 \times 65 \times 65$, so that the spherical grid is of size $64 \times 32 \times 65$, this will allow us to extract spherical images of size 64×64 similar to Esteves et al. [18] and Mukhaimar et al. [21]. It takes about 2.3 s to compute SPFT per volume/object and these are computed for all 12 311 objects for all cases. Table 3 shows the classification results as compared to all the other methods discussed for ModelNet40 dataset.

For our classification task, we did not significantly modify the spherical convolution network proposed in [21]. We adapted the input data to use the magnitude spectrum data at various radii obtained after computing SPFT and kept the number of radii used for concentric spheres (spatial for Mukhaimar et al. [21] while spectral for our solution) the same as 7. We also altered some hyper parameters, such as increasing the filter order for SH to 13 instead of 9, which we found suitable and useful based on our experiments. The reasoning for this was that the spherical images from spatial occupancy grid were sparse, but the spherical images from SPFT were, besides having other characteristics, dense and smooth. However, unlike Ayman et al. [21], we do not include the azimuthal rotations within our training set, as this network is not necessarily **SO**(**3**) group equivariant. For all other methods, we simply follow the training and the test procedures as prescribed in their papers, which includes random azimuthal rotations

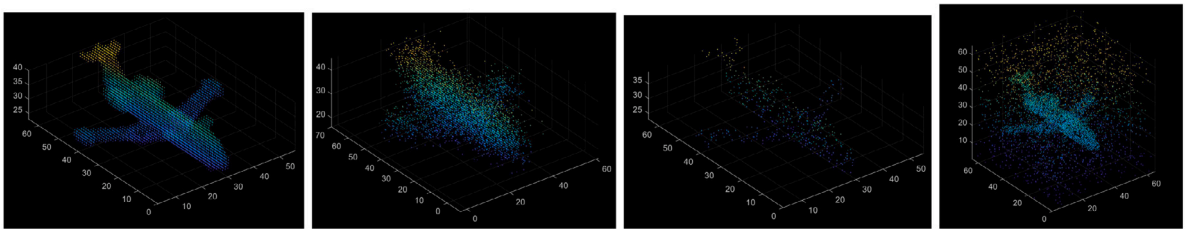

Fig. 27. Four scenarios shown with a model of airplane for training and test, clean, noisy, missing data and outliers.

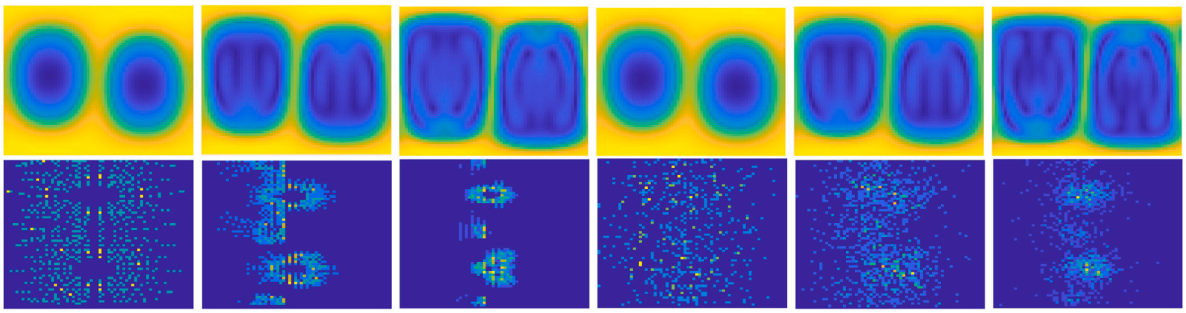

Fig. 28. For the airplane model used in Fig. 27, top row, first half – SPFT based spherical images at radii, $r = 1, 2, 3$ – clean, next three images are still SPFT based but infected with noise. Second row, first half — spatial occupancy grid based spherical images used in [21], next three images are also using same method but are infected with noise. We note that SPFT based image features, especially the lower order radii or low frequency signals, in the spectral domain are very robust to noise as compared to the spatial domain spherical images.

Table 3
The classification accuracy on ModelNet40 dataset against different types of noise, perturbations and corruptions.

Method	Clean	Outliers	Noise	Missing
Spherical-CNN[a] [18]	87	5	53	27
PointNet[b] [52]	89	4	27	57
KPConv[b] [56]	89	4	4	12
SphNet[b] [53]	91	5	5	5
Dynamic Graph CNN[b] [54]	**92.2**	5	5	18
PL-Net3d[b] [55]	86.6	70	60	50
OctNet[c] [50]	86	63	55	25
R-SCNN[c] [21]	82.2	72	63	58
SPFT+[c] **(ours)**	80.63	**78.77**	**79.82**	**76.94**
VoxNet[c] [48]	86	7	17	28

[a] Methods uses input as spherical images transformed using a single sphere surface.
[b] Methods uses directly 3D points/point clouds.
[c] Methods uses voxels or perfect Euclidean grids, either transformed from point cloud or maybe natively present in this form such as in medical imaging, as an input to the respective methods.

and small jittering. When we do include the azimuthal rotations within our training set, the accuracy using our method drops in all cases, to about 70%, confirming that the network is not well designed to be considered group equivariant. It is also interesting to note that OctNet, Riegler et al. [50], which uses 3D convolutional networks, that are both deep and high resolution, is inherently robust to all types of noise in comparison, while the other two recent works, Mukhaimar et al. [21,55], were specifically designed to achieve this type of insensitivity to such perturbations and corruptions.

Fig. 28 shows a side-by-side comparison of the two types of input spherical images used here, one for the SPFT and other based on spatial occupancy grid. This spherical convolution network, both the original and a slightly modified version that we use, has a uniform filter order for each branch of concentric sphere, as it was designed for spatial occupancy grid type of data. SPFT based spherical images have more structure, as they are ordered from low to high frequencies with increasing details. Hence, the network design could have variable filter order of SH at different radii. We note that the spatial occupancy grid or other spatial projection based images are sparse and scattered, that tear or spread the volumetric data, while the images from SPFT are smooth and symmetrical, preserving the structure in an organized way.

Note also that we have not experimented sufficiently with ground up designs of possible different spherical CNN architectures that are truly capable of leveraging all the features of special Fourier domain data discussed here, such as smoothness, structure, symmetry, redundancy and spherical harmonics based possible rotational equivariance. In our classification experiments we have also only used the spectral magnitude information exclusively and not combined it with spatial spherical information also available because before that some design questions would need to be answered. Is it better to just combine all channels together, regardless of the domain, as an input to the network with single/shared weights? Or to have separate voting networks or branches that eventually combine at the end and allow for better empirical results? There are real difficulties in design of a good network architecture that usually requires expert knowledge/input, analysis and understanding. Esteves et al. [18] uses a single surface sphere, while Mukhaimar et al. [21] uses sampling over concentric spheres to mitigate the effects of noise and hence to achieve some degree of robustness. But note that both are still using spatial data only. We found through our experiments, that spherical CNN proposed in [18], even though implemented and argued as using multi-valued/multi-channeled spherical functions to achieve rotational equivariance, it does not simply work by extending its channels directly from two to more than two and using spherical functions from SPFT. Their main architecture has two branches, one for distances and one for surface normal. The two channels used in it encode spatial data, and we found that in our experiments that it is not easily extensible for a multi-radii spectral spherical inputs. This network is deep and well designed to be rotational equivariant, but it has poor mitigation of noise and other forms of perturbations, as also discussed in [21]. Their architecture however, though not group equivariant and hence suffering from sufficient rotational perturbations, and also not as deep as Esteves et al. [18], was amenable to multi-radii spectrum data as an input and easily extensible, hence we adopted it for our demonstration using SPFT.

The above experiments clearly show that spherical images computed from SPFT can be indeed robust and reliable spectral signatures of the original spatial 3D objects. They can be effectively used

for estimating rotations using spherical correlations, SOFT [45], or for classification using spherical convolutions with Spherical CNNs networks such as Esteves et al. [18] and Mukhaimar et al. [21]. Further, this type of Fourier analysis is indeed possible directly on non-Euclidean data, Bronstein et al. [51], such as point cloud, where a set of non-uniformly sampled points in space can be transformed using non-uniform fast Fourier transforms [70].

Our primary goal here is not to argue that our methods are the best performing or are state-of-the-art, but instead to introduce, describe and demonstrate that the spectral spherical Fourier data and the related signal processing tools such as convolutions, corrections and spherical harmonics, and the techniques such as redundancy to achieve robustness, can be effectively put to excellent use in certain fundamental computer vision tasks such as registration and classification. To our knowledge this is the first work that is entirely built using multi-radii spectral spherical data, and also where the whole registration or classification pipelines are performed purely in the Fourier domain and using spherical coordinates. Esteves et al. [18] and Mukhaimar et al. [21], both spherical CNNs, rely only on spatial data as its input, where it requires voxel grid of single or concentric spheres in spatial domain to learn features over the unit ball, before further processing in Fourier domain using spherical harmonics. We also think that the use of both spectral and spatial data, both spherical in nature but as global and local descriptors respectively, the encoder–decoder architecture part or 'what' and 'where' of the CNN literature [71], along with similar Euclidean counterparts, would likely be potentially beneficial for better network designs especially at least in terms of robustness, if possibly also for accuracy. Hopefully our paper may help inspire some research steps in this novel direction. Due to the nature of SPFT, and the separability achieved in rotational (magnitude) and translational (phase) parts, a network could be well designed that can be simultaneously rotational and translational equivariant.

5. Contributions and conclusions

The innovation in our proposed method for both robust 3D registration and learning is the use of multi-radii spectral data for spherical correlation/convolution on the spheres, which have the redundant 3D rotational information encoded across all spherical layers, to perform the globally optimal 3D rotation estimation, best selected from a set of solutions present at different radii, and obtained with the combination of fast and algebraically exact SPFT, and SOFT algorithms, or learned using spherical CNNs.

The main contributions of our paper are summarized:

- A general method of registration given the two 3D volumes in Cartesian coordinates related by up to 7 DOF transformation.
- A novel approach of using spherical spectral signatures to train spherical convolutional networks.
- It is extremely robust to noise, partial overlap or missing data, and other distortions, due to the redundant nature of the computations using multi-layered spectrum.
- The registration algorithm presented provides a robust global estimate, using multi-radii SPFT for the rotation estimation problem, often considered as the most difficult part of the 3D registration process.
- The robustness achieved in 3D learning classification task, in the presence of significant corruptions, also follows the same exact fundamental principle as the registration algorithm, which is through the well designed leveraging of the multi-radii SPFT.
- For volumetric registration the component pieces of the pipeline, SPFT and SOFT, are both of fixed complexities, and in principle a parallel implementation can take full advantage of commodity GPGPU processing hardware. Similarly for the machine learning task, the tools such as spherical convolutions and spherical harmonics. Most importantly, each of these have a parallelizable structure and scales naturally with processing and memory resources.
- We also demonstrate through a robust classification machine learning task, that the spectral spherical signatures, obtained using SPFT, are robust, reliable and effective against various types of real world noise, data corruptions and perturbations.
- Finally, our registration solution is easily extensible to higher dimensions, again using fast and exact FT computations for *hyper-spherical* FT and a similar *hyper* SOFT, using Abbas et al. [44] and Kostelec and Rockmore [45], where the geometric structure is highly rich and complex.

Further research is needed to seek more problems where the proposed registration and learning technique, along with all the signal processing tools discussed can be successfully applied. Also there exists no deep learning network in the literature that is both robust to noise, perturbations, outliers while remaining group equivariant to arbitrary rotations in the input stage, or how this type of information can be integrated into existing architectures, or even constructed from first principles, but it would be interesting to explore in these research directions. A potential application would also be to design of networks that leverage simultaneously both spatial and spectral domain, along with Euclidean and spherical data, in a combined and hybrid way. It would be also beneficial to study how it can be extended for more complicated transformations such as affine and projective. In future, it would be very interesting to see an application of high dimensional extension of the proposed technique in terms of performance, complexity, parallel implementation, network constraints, etc.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Source code and some supplementary materials will be made openly available at: https://github.com/syedalamabbas/Robust-Registration-Learning-SPFT.

Acknowledgments

We would like to thank Drs. M.H. Mahoor and H. Bulow, and their collaborators for providing the 3D rotation estimation *MATLAB* code and datasets for their paper, which along with the publicly available code of Go-ICP by Dr. Jiaolong helped us to do the comparisons. We also acknowledge the publicly available C++ SOFT package by Dr. P. Kostelec that was used in our research. We thank Ayman et al. for making R-SCNN code available on GitHub that was used in our study, and Vinay B. R. Guduri for helping in our machine learning setup.

References

[1] J. Le Moigne, Introduction to remote sensing image registration, in: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 2017, pp. 2565–2568.
[2] M. Druckmüller, Phase correlation method for the alignment of total solar eclipse images, Astrophys. J. 706 (2) (2009) 1605.
[3] D. Hong, N. Yokoya, J. Chanussot, X.X. Zhu, An augmented linear mixing model to address spectral variability for hyperspectral unmixing, IEEE Trans. Image Process. 28 (4) (2018) 1923–1938.
[4] D. Çeşmeci, M. Güllü, Phase-correlation-based hyperspectral image classification using multiple class representatives obtained with k-means clustering, Int. J. Remote Sens. 30 (14) (2009) 3827–3834.
[5] D. Scheffler, A. Hollstein, H. Diedrich, K. Segl, P. Hostert, AROSICS: An automated and robust open-source image co-registration software for multi-sensor satellite data, Remote Sens. 9 (7) (2017) 676.
[6] C.A. Wilson, J.A. Theriot, A correlation-based approach to calculate rotation and translation of moving cells, IEEE Trans. Image Process. 15 (7) (2006) 1939–1951.

[7] D.J. Kelly, E.U. Azeloglu, P.V. Kochupura, G.S. Sharma, G.R. Gaudette, Accuracy and reproducibility of a subpixel extended phase correlation method to determine micron level displacements in the heart, Med. Eng. Phys. 29 (1) (2007) 154–162.
[8] D. Morley, H. Foroosh, Computing cyclotorsion in refractive cataract surgery, IEEE Trans. Biomed. Eng. 63 (10) (2015) 2155–2168.
[9] M.S.M. Asaari, S.A. Suandi, B.A. Rosdi, Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics, Expert Syst. Appl. 41 (7) (2014) 3367–3382.
[10] I. Rida, S. Almaadeed, A. Bouridane, Gait recognition based on modified phase-only correlation, Signal Image Video Process. 10 (2016) 463–470.
[11] A. Kumar, Toward more accurate matching of contactless palmprint images under less constrained environments, IEEE Trans. Inf. Forensics Secur. 14 (1) (2018) 34–47.
[12] H. Bülow, A. Birk, Spectral registration of noisy sonar data for underwater 3D mapping, Auton. Robots 30 (2011) 307–331.
[13] N. Hurtos, D. Ribas, X. Cufí, Y. Petillot, J. Salvi, Fourier-based registration for robust forward-looking sonar mosaicing in low-visibility underwater environments, J. Field Robotics 32 (1) (2015) 123–151.
[14] H. Moriya, Phase-only correlation of time-varying spectral representations of microseismic data for identification of similar seismic events, Geophysics 76 (6) (2011) WC37–WC45.
[15] G. Tomar, S.C. Singh, J.-P. Montagner, Sub-sample time shift and horizontal displacement measurements using phase-correlation method in time-lapse seismic, Geophys. Prospect. 65 (2) (2017) 407–425.
[16] X. Tong, Z. Ye, Y. Xu, S. Gao, H. Xie, Q. Du, S. Liu, X. Xu, S. Liu, K. Luan, et al., Image registration with Fourier-based image correlation: A comprehensive review of developments and applications, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (10) (2019) 4062–4081.
[17] D.E. Worrall, S.J. Garbin, D. Turmukhambetov, G.J. Brostow, Harmonic networks: Deep translation and rotation equivariance, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5028–5037.
[18] C. Esteves, C. Allen-Blanchette, A. Makadia, K. Daniilidis, Learning so (3) equivariant representations with spherical cnns, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 52–68.
[19] T.S. Cohen, M. Geiger, J. Köhler, M. Welling, Spherical cnns, 2018, arXiv preprint arXiv:1801.10130.
[20] T. Cohen, M. Welling, Group equivariant convolutional networks, in: International Conference on Machine Learning, PMLR, 2016, pp. 2990–2999.
[21] A. Mukhaimar, R. Tennakoon, C.Y. Lai, R. Hoseinnezhad, A. Bab-Hadiashar, Robust object classification approach using spherical harmonics, IEEE Access 10 (2022) 21541–21553.
[22] R.N. Bracewell, R.N. Bracewell, The Fourier Transform and Its Applications, Vol. 31999, McGraw-Hill New York, 1986.
[23] H. Shekarforoush, M. Berthod, J. Zerubia, Subpixel image registration by estimating the polyphase decomposition of cross power spectrum, in: Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 1996, pp. 532–537.
[24] B.S. Reddy, B.N. Chatterji, An FFT-based technique for translation, rotation, and scale-invariant image registration, IEEE Trans. Image Process. 5 (8) (1996) 1266–1271.
[25] H. Foroosh, J.B. Zerubia, M. Berthod, Extension of phase correlation to subpixel registration, IEEE Trans. Image Process. 11 (3) (2002) 188–200.
[26] M. Balci, H. Foroosh, Subpixel estimation of shifts directly in the Fourier domain, IEEE Trans. Image Process. 15 (7) (2006) 1965–1972.
[27] M. Guizar-Sicairos, S.T. Thurman, J.R. Fienup, Efficient subpixel image registration algorithms, Opt. Lett. 33 (2) (2008) 156–158.
[28] G. Tzimiropoulos, V. Argyriou, T. Stathaki, Subpixel registration with gradient correlation, IEEE Trans. Image Process. 20 (6) (2010) 1761–1767.
[29] B. Sun, W. Kong, L. Zhang, J. Zhang, Fourier analysis techniques applied in data registration: A survey, in: 2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI), 2014, pp. 1–5, http://dx.doi.org/10.1109/MFI.2014.6997740.
[30] S. Althloothi, M.H. Mahoor, R.M. Voyles, A robust method for rotation estimation using spherical harmonics representation, IEEE Trans. Image Process. 22 (6) (2013) 2306–2316.
[31] H. Bülow, A. Birk, Spectral 6DOF registration of noisy 3D range data with partial overlap, IEEE Trans. Pattern Anal. Mach. Intell. 35 (4) (2012) 954–969.
[32] P.J. Besl, N.D. McKay, Method for registration of 3-D shapes, in: Sensor Fusion IV: Control Paradigms and Data Structures, Vol. 1611, SPIE, 1992, pp. 586–606.
[33] J. Yang, H. Li, Y. Jia, Go-icp: Solving 3d registration efficiently and globally optimally, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 1457–1464.
[34] J. Yang, H. Li, D. Campbell, Y. Jia, Go-ICP: A globally optimal solution to 3D ICP point-set registration, IEEE Trans. Pattern Anal. Mach. Intell. 38 (11) (2015) 2241–2254.
[35] R.B. Rusu, S. Cousins, 3D is here: Point cloud library (PCL), in: 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 1–4, http://dx.doi.org/10.1109/ICRA.2011.5980567.
[36] D. Holz, A.E. Ichim, F. Tombari, R.B. Rusu, S. Behnke, Registration with the point cloud library: A modular framework for aligning in 3-D, IEEE Robot. Autom. Mag. 22 (4) (2015) 110–124, http://dx.doi.org/10.1109/MRA.2015.2432331.
[37] C. Wang, X. Jing, C. Zhao, Local upsampling Fourier transform for accurate 2D/3D image registration, Comput. Electr. Eng. 38 (5) (2012) 1346–1357, http://dx.doi.org/10.1016/j.compeleceng.2012.04.005, URL: https://www.sciencedirect.com/science/article/pii/S0045790612000778. Special issue on Recent Advances in Security and Privacy in Distributed Communications and Image processing.
[38] J. Luce, J. Gray, M.A. Hoggarth, J. Lin, E. Loo, M.I. Campana, J.C. Roeske, Medical image registration using the Fourier transform, Int. J. Med. Phys. Clin. Eng. Radiat. Oncol. 2014 (2014).
[39] L. Lucchese, G. Doretto, G.M. Cortelazzo, A frequency domain technique for range data registration, IEEE Trans. Pattern Anal. Mach. Intell. 24 (11) (2002) 1468–1484.
[40] A. Bermanis, A. Averbuch, Y. Keller, 3-d symmetry detection and analysis using the pseudo-polar fourier transform, Int. J. Comput. Vis. 90 (2010) 166–182.
[41] Y. Keller, Y. Shkolnisky, A. Averbuch, Volume registration using the 3-D pseudopolar Fourier transform, IEEE Trans. Signal Process. 54 (11) (2006) 4323–4331.
[42] M. Kazhdan, T. Funkhouser, Harmonic 3D shape matching, in: ACM SIGGRAPH 2002 Conference Abstracts and Applications, 2002, pp. 191–191.
[43] A. Averbuch, Y. Shkolnisky, 3D Fourier based discrete radon transform, Appl. Comput. Harmon. Anal. 15 (1) (2003) 33–69.
[44] S.A. Abbas, Q. Sun, H. Foroosh, An exact and fast computation of discrete Fourier transform for polar and spherical grid, IEEE Trans. Signal Process. 65 (8) (2016) 2033–2048.
[45] P.J. Kostelec, D.N. Rockmore, FFTs on the rotation group, J. Fourier Anal. Appl. 14 (2008) 145–179.
[46] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz, Rotation invariant spherical harmonic representation of 3 d shape descriptors, in: Symposium on Geometry Processing, Vol. 6, 2003, pp. 156–164.
[47] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, M. Bennamoun, Deep learning for 3d point clouds: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (12) (2020) 4338–4364.
[48] D. Maturana, S. Scherer, Voxnet: A 3d convolutional neural network for real-time object recognition, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2015, pp. 922–928.
[49] D. Stutz, A. Geiger, Learning 3d shape completion from laser scan data with weak supervision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1955–1964.
[50] G. Riegler, A.O. Ulusoy, A. Geiger, OctNet: Learning deep 3D representations at high resolutions, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6620–6629, http://dx.doi.org/10.1109/CVPR.2017.701.
[51] M.M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: going beyond euclidean data, IEEE Signal Process. Mag. 34 (4) (2017) 18–42.
[52] C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652–660.
[53] A. Poulenard, M. Rakotosaona, Y. Ponty, M. Ovsjanikov, Effective rotation-invariant point CNN with spherical harmonics kernels, in: 2019 International Conference on 3D Vision (3DV), 2019, pp. 47–56, http://dx.doi.org/10.1109/3DV.2019.00015.
[54] Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein, J.M. Solomon, Dynamic graph CNN for learning on point clouds 38, 2019, http://dx.doi.org/10.1145/3326362.
[55] A. Mukhaimar, R. Tennakoon, C.Y. Lai, R. Hoseinnezhad, A. Bab-Hadiashar, PL-net3d: Robust 3D object class recognition using geometric models, IEEE Access 7 (2019) 163757–163766, http://dx.doi.org/10.1109/ACCESS.2019.2952638.
[56] H. Thomas, C.R. Qi, J. Deschaud, B. Marcotegui, F. Goulette, L. Guibas, Kpconv: Flexible and deformable convolution for point clouds, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6410–6419, http://dx.doi.org/10.1109/ICCV.2019.00651.
[57] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao, 3D shapenets: A deep representation for volumetric shapes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1912–1920.
[58] Y. Wang, C. Yan, Y. Feng, S. Du, Q. Dai, Y. Gao, Storm: Structure-based overlap matching for partial point cloud registration, IEEE Trans. Pattern Anal. Mach. Intell. 45 (1) (2022) 1135–1149.
[59] A. Kanezaki, Y. Matsushita, Y. Nishida, Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5010–5019.
[60] A. Papoulis, Signal Analysis, Mcgraw-Hill College, 1977.
[61] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.
[62] J.L. Horner, P.D. Gianino, Phase-only matched filtering, Appl. Opt. 23 (6) (1984) 812–816.

[63] P.J. Kostelec, The soft package, 2015, Retrieved from, http://www.cs.dartmouth.edu/~geelong/soft/.

[64] D.J. Rosenman, Chapel Hill Volume Rendering Test Data Set, Volume II, North Carolina Memorial Hospital, University of North Carolina, 2015, Retieved from https://graphics.stanford.edu/data/voldata/.

[65] P. Shilane, P. Min, M. Kazhdan, T. Funkhouser, The princeton shape benchmark, in: Proceedings Shape Modeling Applications, 2004, IEEE, 2004, pp. 167–178.

[66] S. Li, SPHARM-MAT, a Matlab-Based 3D Shape Modeling and Analysis Toolkit, Perelman School of Medicine, University of Pennsylvania, 2017, Retrieved from, https://www.med.upenn.edu/shenlab/spharm-mat.html.

[67] A. Fabri, S. Pion, CGAL: The computational geometry algorithms library, in: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2009, pp. 538–539.

[68] J. Pluim, J. Maintz, M. Viergever, Mutual-information-based registration of medical images: a survey, IEEE Trans. Med. Imaging 22 (8) (2003) 986–1004, http://dx.doi.org/10.1109/TMI.2003.815867.

[69] J. Kim, J. Fessler, Intensity-based image registration using robust correlation coefficients, IEEE Trans. Med. Imaging 23 (11) (2004) 1430–1444, http://dx.doi.org/10.1109/TMI.2004.835313.

[70] J. Keiner, S. Kunis, D. Potts, Using NFFT 3—a software library for various nonequispaced fast Fourier transforms, ACM Trans. Math. Softw. 36 (4) (2009) 1–30.

[71] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967–5976, http://dx.doi.org/10.1109/CVPR.2017.632.