



Consistent Correspondence of Cone-Beam CT Images Using Volume Functional Maps

Yungeng Zhang¹, Yuru Pei^{1(✉)}, Yuke Guo², Gengyu Ma³, Tianmin Xu⁴,
and Hongbin Zha¹

¹ Key Laboratory of Machine Perception (MOE), Department of Machine Intelligence, Peking University, Beijing, China

peiyuru@cis.pku.edu.cn

² Luoyang Institute of Science and Technology, Luoyang, China

³ uSens Inc., San Jose, USA

⁴ School of Stomatology, Peking University, Beijing, China

Abstract. Dense correspondence between Cone-Beam CT (CBCT) images is desirable in clinical orthodontics for both intra-patient treatment evaluation and inter-patient statistical shape modeling and attribute transfer. Conventional 3D deformable image registration relies on time-consuming iterative optimization for correspondences. The recent forest-based correspondence methods often require large offline training costs and a separate regularization in the post-processing. In this work, we propose an efficient volume functional map for dense and consistent correspondence between CBCT images. We design a group of volume functions specifically for CBCT images and construct a reduced functional space on supervoxels. The low-dimensional map between the limited spectral bases determines the dense supervoxel-wise correspondence in an unsupervised way. Further, we perform consistent functional mapping in a collection of volume images to handle ambiguous correspondences of craniofacial structures, e.g., those due to the intercuspation. A subset of orthonormal volume functional maps is optimized on a Stiefel manifold simultaneously, which determines the cycle-consistent pairwise functional maps in the volume collection. Benefits of the proposed volume functional maps have been illustrated in label propagation and segmentation transfer with improved performance over conventional methods.

1 Introduction

Malocclusion has a high prevalence and causes aesthetic and functional problems in a large population. Cone-Beam CT (CBCT) images are widely used in clinical orthodontics to provide joint 3D geometries of the teeth, mandible, and maxilla to facilitate accurate malocclusion diagnoses and treatment evaluations. Efficient dense correspondence between CBCT images is desirable in several scenarios, including measuring shape variations due to treatments and growth [8], label propagation [5], and statistical craniofacial shape modeling [6].

The conventional 3D deformable registration methods, such as B-spline and Demons based registrations [11], solve the dense displacement field and correspondence under a time-consuming large-scale optimization for the craniofacial CBCT images. The importance sampling is helpful to accelerate the deformable registration by reducing the parameter space with efficient Jacobian estimation of similarity metrics [2]. However, the registration of the reduced subset also relies on the online iterative optimization. The random forest realizes efficient online dense correspondences between 3D surface meshes [10] and volume images [5, 9]. Aside from the supervised classification random forest learned from a large set of labeled 3D meshes [10] or pseudo labeling obtained by supervoxel decompositions [5], the unsupervised clustering random forest realizes the self-learning of data distribution and affinity estimation without prior labeling [9]. However, the random forest built on independent data points could not guarantee the spatial consistency. A separate regularization scheme is required for smooth correspondences [9, 10]. Recently the spectral methods using the Laplace Beltrami operator have gained popularity for functional mapping [7, 14], co-segmentation [12], and analysis of anatomical structure [6] on surfaces and images. The functional map has a high efficiency by performing spectral mapping in a reduced functional space. However, the previous functional maps only handle 2D manifolds including images [12] and 3D surfaces [6, 7].

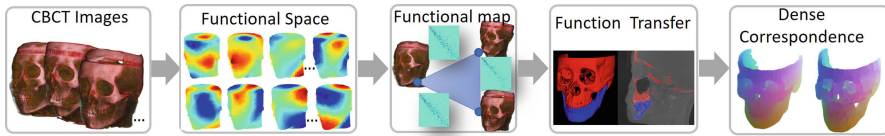


Fig. 1. Flowchart of our system.

In this paper, we propose a novel volume functional map for establishing supervoxel-wise correspondences between CBCT images (see Fig. 1). The proposed method extends the existing functional map approach from a 2D manifold, including the 2D image and 3D surface, to the 3D volume image. We design a group of volume functions, including appearances, contexts, geodesics, and label maps on supervoxel specifically for consistent correspondences between CBCT images. The spectral decomposition of the graph Laplacian produces harmonics bases of each volume image to span a linear volume functional space. The scalar-valued functions of both features and attributes over supervoxels can be reconstructed from a reduced set of functional bases. The dense supervoxel-wise correspondence is realized by finding a spectral transformation matrix between reduced functional spaces. The functional map is optimized by aligning the volume functions in an unsupervised way. Furthermore, in order to reduce correspondence ambiguities of craniofacial structures, e.g., the separation of upper and lower dentitions due to intercuspatation, we exploit the cycle consistency constraints by introducing a latent functional space to a volume collection.

The pairwise orthonormal functional maps in the volume collection are optimized simultaneously on a Stiefel manifold, which meet the invertibility and transitivity requirements. The volume functional map realizes online label propagation and attribute transfer between volume images by the linear algebra with less computational complexity than conventional methods.

2 Methods

The input is a collection of clinically captured craniofacial CBCT images $\mathcal{V} = \{V_1, \dots, V_N\}$. The goal is to build dense supervoxel-wise correspondences between volume images. Without loss of generality, we decompose each volume image into supervoxels. A volume image is represented by a graph $\mathcal{G} = (S, \mathcal{E})$ over the supervoxels $S = \{s_i | i = 1, \dots, M\}$. \mathcal{E} denotes the edges connecting adjacent supervoxels, which are weighted according to the affinity of adjacent supervoxels. In the unsupervised setting, the supervoxel-wise mapping $P_{ij} \in \mathbb{R}^{M \times M}$ between image V_i and V_j is solved based on the alignment of multi-channel features. The system also allows a user to label a small set of landmarks or region correspondences in a semi-supervised setting. With this setup, the goal is to estimate a permutation matrix P_{ij} of all supervoxels regarding CBCT images V_i and V_j .

Volume Functions. In our system, both features and attributes of supervoxels are represented by real-valued functions. Denote function $f : S \rightarrow \mathbb{R}$ to map a supervoxel s to a real value $g(s) \in \mathbb{R}$. There are four types of functions regarding the supervoxel appearance, context, geodesic distance, and label maps. The first three types are continuous real-valued functions, whereas the last one is a binary function. The appearance functions of supervoxels are composed of the normalized histograms of the original intensity and intensity gradients in x, y , and z directions. The context functions are composed of appearance differences of one supervoxel to those in a predefined contextual pattern [9]. The geodesic distance functions are defined by the sorted distance vector $\kappa(d_{i',j'} | j' = 1, \dots, M_*)$ between supervoxel $s_{i'}$ to the rest supervoxels on the weighted graph \mathcal{G} , where $d_{i',j'}$ is the shortest graph distance between supervoxel $s_{i'}$ and $s_{j'}$. κ is a cubic-spline fitting and resampling operator on the sorted distance vector. In our system, we only compute geodesic vectors of M_* bony supervoxels for the computational efficiency. The label maps defined by a user are only used in the semi-supervised setting, where the indicator function $g(s) = 1$ for corresponding landmarks or regions, and $g(s) = 0$ otherwise. Let \mathbb{G}_i denote all volume functions over supervoxels of image V_i . The functions \mathbb{G}_i spans a linear space in \mathbb{R}^M .

Reduced Volume Functional Space. The Laplace-Beltrami operator on a manifold is defined as the divergence of the gradient, $\Delta g = \text{div} \nabla g$. The eigendecomposition, $\Delta \phi = \lambda \phi$, results in harmonic bases of the functional space with frequencies λ . On the discrete supervoxel decomposed volume image, the graph Laplace is used to approximate the Laplace Beltrami operator. Let W denote the weighted adjacency matrix of supervoxel graph \mathcal{G} , $L = D^{-1}(D - W)$, where $D_{ii} = \sum_j W_{ij}$. The eigendecomposition of L results in eigenvectors

$\Phi = (\phi_1, \dots, \phi_M)$ as the harmonics bases and eigenvalues $(\lambda_1, \dots, \lambda_M)$ as harmonics frequencies. The eigenvectors are sorted according to the harmonic frequencies, and the first K eigenvectors are used to represent the reduced functional space. K is set at 75 in our experiments. Eight eigenvectors related to a volume image are illustrated in Fig. 2(a). The volume function is represented as a linear combination of eigenvectors, $g = \Phi \mathbf{g}$, where $\mathbf{g} \in \mathbb{R}^K$. The reduced bases $\Phi^* \in \mathbb{R}^{M \times K}$.

2.1 Volume Functional Map

Given a volume image pair (V_i, V_j) , and a volume function $g^{(i)} = \Phi^{(i)} \mathbf{g}^{(i)} \in \mathbb{G}_i$, the goal of volume functional mapping is to transfer the K -dimensional vector $\mathbf{g}^{(i)}$ to the functional space of image V_j , and reconstruct the volume function $g^{(j)} \in \mathbb{G}^{(j)}$. Given H corresponding functions $\mathbb{G}_i \in \mathbb{R}^{M \times H}$ and $\mathbb{G}_j \in \mathbb{R}^{M \times H}$ on image V_i and V_j , the corresponding supervoxels between volume images should have similar functional values. The objective function $E = \|\mathbb{G}_i - P\mathbb{G}_j\|_F^2$, where P is the unknown permutation matrix indicating the dense supervoxel correspondence between V_i and V_j . Instead of the supervoxel-wise correspondence, we handle the low-dimensional functional map C_{ij} between the reduced functional spaces. The functional map $C_{ij} = \Phi^{(i)-1} P_{ij} \Phi^{(j)}$ [6, 7]. The transferred function $g^{(j)} = \Phi^{(j)} C_{ij} \mathbf{g}^{(i)}$. The functional map is viewed as a spectral transformation of the reduced functional space $\Phi^{(i)}$ and $\Phi^{(j)}$, in which the transformation matrix accounts for the sign flipping and interchanging of eigenvectors between volume images. It is straightforward that the functional map between image V_i and V_j should transform the feature function $g^{(i)} \in \mathbb{G}_i$ to the feature function $g^{(j)} \in \mathbb{G}_j$. The functional map is optimized by minimizing feature alignment errors.

$$E(C_{ij}) = \|C_{ij} \overline{\mathbf{g}^{(i)}} - \overline{\mathbf{g}^{(j)}}\|_F^2 + \gamma \|\Theta_j C_{ij} - C_{ij} \Theta_i\|_F^2, \quad (1)$$

where $\|\cdot\|_F$ is the Frobenius norm. $\overline{\mathbf{g}} \in \mathbb{R}^{K \times H}$ denotes the harmonic weight matrix in the reduced functional space. The feature space of image V_i is aligned to that of image V_j by minimizing the first term. The second term is the operator commutativity constraints. Θ is a low rank approximation of the Graph Lapidarian matrix. The constant γ is used to balance the feature alignment and the commutativity constraint, and set at 1 in our experiments. We use the linear least square to solve C_{ij} . Given functional map C_{ij} , the dense correspondence matrix $P_{ij}^* = \Phi^{(i)} C_{ij} \Phi^{(j)-1}$. Note that the matrix P_{ij}^* is not a hard permutation between image V_i and V_j , since the entries record the probability of supervoxel pair (s_i, s_j) being a counterpart to each other. The permutation matrix P_{ij} is derived from P_{ij}^* by using the column normalization and the NN scheme [7].

Consistency Regularization. When given additional images, cycle-consistent functional maps in an image collection are helpful to improve the mapping accuracies over the pairwise functional maps [12, 14]. In our system, we utilize the consistency regularization to reduce the mapping ambiguity especially for the

segmentation transfer of the mandible and maxilla. We follow the map decomposition [12], where the functional maps, $C_{i,j} = \mathbf{c}_j' \mathbf{c}_i$, are determined by a reduced mapping set $\{\mathbf{c}_1, \dots, \mathbf{c}_M\}$. \mathbf{c}_i can be viewed as the functional map from reduced functional space of V_i to a latent functional space. The decomposition of $C_{i,j}$ enforces the 3-cycle consistency of the functional maps in a volume collection. We further require the functional map $\mathbf{c}_i, 1 \leq i \leq M$, be an orthonormal matrix in the Stiefel manifold. Thus, all the functional maps are orthonormal, and $C_{ij}' = C_{ij}^{-1}$. The functional maps satisfies the invertibility and transitivity constraints, where $C_{ij} = C_{ji}^{-1}$ and $C_{jk}C_{ij} = C_{ik}$. The objective function is rewritten as

$$E(\mathbf{c}) = \sum_{V_i, V_j \in \mathcal{V}, C_{i,j} = \mathbf{c}_j' \mathbf{c}_i} \|C_{ij} \overline{\mathbf{g}}^{(i)} - \overline{\mathbf{g}}^{(j)}\|_F^2 + \gamma \|\Theta_j C_{ij} - C_{ij} \Theta_i\|_F^2. \quad (2)$$

We implement the optimization of the functional map \mathbf{c} on the Stiefel manifold using the trust region solver of the Manopt toolbox [3]. The functional maps \mathbf{c} are initialized as an identity matrix and refined using the manifold optimization. In the online testing, the corresponding volume functions are extracted from the novel CBCT image. The pairwise volume functions map is computed by minimizing Eq. 1. When given additional volume images, the consistent volume functional maps are obtained by minimizing Eq. 2.

3 Experiments

Dataset. The proposed volume functional map is evaluated on a collection consisting of 10 clinically captured CBCT images of orthodontic patients, which has 90 pairwise maps. The volume image is of a resolution of $250 \times 250 \times 238$ with a voxel size of $0.8 \text{ mm} \times 0.8 \text{ mm} \times 0.8 \text{ mm}$. We use the SLIC method [1] to decompose each CBCT image into $20k$ supervoxels. For each CBCT image, there are 680 functions, including 80 appearances, 500 contexts, 100 geodesics-related functions.

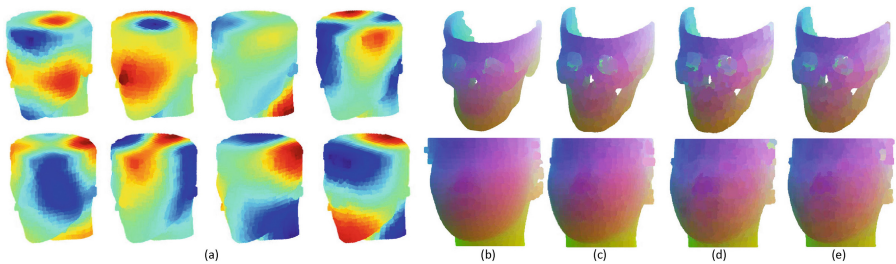


Fig. 2. (a) Eight functional bases. Supervoxel-wise correspondence between (b) the reference image and the target image by (c) the deformable B-spline registration, and the proposed (d) VFM and (e) C-VFM methods.

Qualitative Assessment. We qualitatively evaluate the supervoxel-wise label propagation of the mandible and maxilla using two metrics: the Dice similarity coefficient (DSC) and the average Hausdorff distance (AHD). We compare the proposed pairwise volume functional map (VFM) and the consistent volume functional map (C-VFM) with the conventional label propagation methods, including the patch fusion (PF) [4], the convex optimization (CO) [13], the volumetric deformable B-spline registration [11]. We also compare with the random forest-based methods, including the classification forest (Cla) [5] and the mixed metric forest (MMRF) [9] as shown in Figs. 2(b-e) and 3(a). The label propagation accuracies of the proposed method have DSCs of 0.94 ± 0.02 and 0.93 ± 0.02 when using 75 spectral bases for the mandible and the maxilla respectively, which are close to the conventional deformable B-spline registration. Moreover, the proposed volume functional map gains great efficiency and consumes approx. 20 s (1.35 s for map optimization (Eq. 1) as shown in Fig. 3(d)) when using a 75×75 functional map vs. 11 min by the B-spline registration with a $28 \times 28 \times 27$ control grid for the segmentation transfer. The running time is measured on a PC with an i7 CPU of 3.3 GHz and RAM of 32GB. The reason for the online efficiency is that the functional map exploits a low dimensional spectral transformation in the reduced function spaces. The volume functional map with a DSC of 0.94 for the mandible label propagation improves over the supervised Cla of 0.88 and the unsupervised MMRF of 0.92. The functional map and the forest-based method both realize efficient online supervoxel-based correspondences, whereas the latter requires a separate regularization and a large offline forest training cost. One sampled functional map is shown in Fig. 3(e).

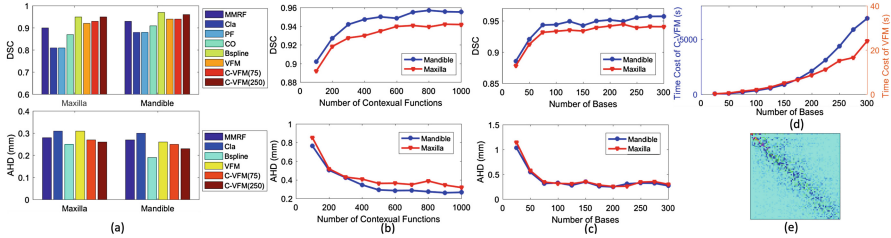


Fig. 3. (a) DSCs and AHDs of the label propagation of the mandible and maxilla by the proposed VFM and C-VFM using 75 and 250 bases compared with PF [4], CO [13], B-spline [11], Cla [5], and MMRF [9] based methods. (b) DSCs and AHDs of the label propagation with increasing number of (b) contextual functions and (c) bases. (d) Time costs of map optimizations of VFM and C-VFM. (e) Functional map of C-VFM.

Since the upper and lower dentitions are assigned to the mandible and maxilla respectively, the intercuspsation causes correspondence ambiguities in segmentation transfer as shown in Fig. 4. The consistency regularization (Sect. 2.1) exploits additional volumes for consistent correspondences. In our experiments, we solve the correspondences between three volumes simultaneously.

The additional volume is helpful to avoid correspondence ambiguities (Fig. 4(f)). Furthermore, the proposed methods can work in a semi-supervised setting, where a user interactively labeled five corresponding landmarks as shown in Fig. 4(c). Corresponding landmarks are represented by pairs of volume functions as described in Sect. 2, and improve the matching even when using a small set of bases.

The functional maps are solved based on the predefined volume functions including the context and geodesic functions. Figure 3(b) illustrates that the label propagation accuracies are positively associated with the number of contextual functions. The geodesic functions facilitate the detection of connected structures. For instance, the geodesic distance between two supervoxels of the same structure is smaller than that of distinct structures. We observe that the functional maps with the geodesic functions are superior to those without the geodesic functions as constraints with mean DSC improvements of 0.53% and 0.56% for the mandible and maxilla respectively.

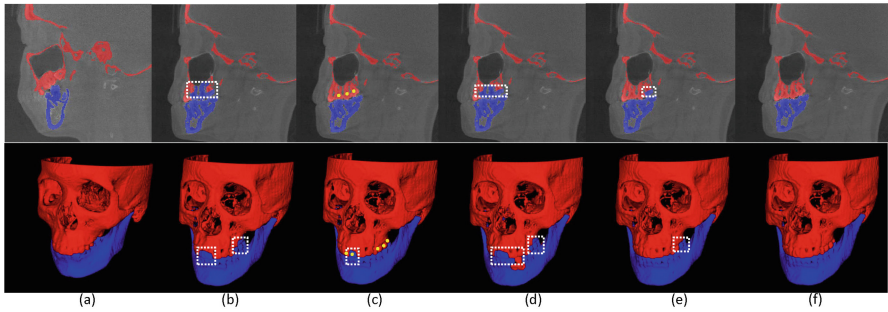


Fig. 4. Segmentation transfer from (a) the reference to the target image using VFM with (b) 25 spectral bases, (c) 25 spectral bases and 5 pairs of landmarks (yellow points), (d) 100 contextual functions, and (e) 75 bases and 500 contextual functions without consistency constraints and (f) with consistency constraints.

In our system, the reduced harmonic bases represent the original functional space compactly. Figure 3(c) shows DSCs of the label propagation with increasing number of harmonic bases. Note that, the more bases used, the more accurate label propagation. For instance, the DSC of the mandible label propagation increases from 0.94 using 75 bases to 0.96 using 250 bases. However, the additional spectral bases increase the computational costs as shown in Fig. 3(d), in which the functional map consumes from 37s to 6850s when using from 25 to 300 spectral bases regarding the C-VFM method.

4 Discussion and Conclusion

In this paper, we extend the conventional functional map on a 2D manifold of surfaces or images to 3D volumes. We propose a novel volume functional map for

supervoxel-wise correspondences between CBCT images for label propagation. The low-dimensional functional map between reduced functional spaces realizes a spectral transformation, and uniquely determines the dense supervoxel correspondence between CBCT images. The proposed consistent volume map is promising to reduce correspondence ambiguities of craniofacial structures, such as those due to the intercuspation. The proposed method has been applied to clinically captured CBCT images for segmentation transfer of the mandible and maxilla with mean DSCs of 0.94 and 0.93 respectively when using 75 spectral bases. However, we observe that volume functional maps are limited to estimate correspondence between volumes with non-isometric deformations, e.g., the volumes of an adult and a child, due to the scale-sensitive context and geodesic functions. In the future work, we would investigate the volume functional map for more general deformations.

Acknowledgments. This work was supported by NSFC 61272342, 61632003, 81371192, ISTCPC 2014DFA31800, and NKTRDPC 2017YFB1002601.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. PAMI* **34**(11), 2274–2282 (2012)
2. Bhagalia, R., Fessler, J.A., Kim, B.: Accelerated nonrigid intensity-based image registration using importance sampling. *IEEE Trans. MI* **28**(8), 1208–1216 (2009)
3. Boumal, N., Mishra, B., Absil, P.A., Sepulchre, R.: Manopt, a matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.* **15**(1), 1455–1459 (2014)
4. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *NeuroImage* **54**(2), 940–954 (2011)
5. Kanavati, F., Tong, T., Misawa, K., Fujiwara, M., Mori, K., Rueckert, D., Glocker, B.: Supervoxel classification forests for estimating pairwise image correspondences. *Pattern Recogn.* **63**, 561–569 (2017)
6. Lombaert, H., Arcaro, M., Ayache, N.: Brain transfer: spectral analysis of cortical surfaces and functional maps. In: Ourselin, S., Alexander, D.C., Westin, C.-F., Cardoso, M.J. (eds.) *IPMI 2015*. LNCS, vol. 9123, pp. 474–487. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19992-4_37
7. Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., Guibas, L.: Functional maps: a flexible representation of maps between shapes. *ACM Trans. Graph.* **31**(4), 30 (2012)
8. Pei, Y., Ma, G., Chen, G., Zhang, X., Xu, T., Zha, H.: Superimposition of cone-beam computed tomography images by joint embedding. *IEEE Trans. BME* **64**(6), 1218–1227 (2017)
9. Pei, Y., et al.: Mixed metric random forest for dense correspondence of cone-beam computed tomography images. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017*. LNCS, vol. 10433, pp. 283–290. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_33
10. Rodolà, E., Rota Bulo, S., Windheuser, T., Vestner, M., Cremers, D.: Dense non-rigid shape correspondence using random forests. In: *CVPR*, pp. 4177–4184 (2014)

11. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: a survey. *IEEE Trans. MI* **32**(7), 1153–1190 (2013)
12. Wang, F., Huang, Q., Guibas, L.J.: Image co-segmentation via consistent functional maps. In: *ICCV*, pp. 849–856 (2013)
13. Wang, L., et al.: Automated segmentation of CBCT image using spiral CT atlases and convex optimization. In: *MICCAI 2013*, pp. 251–258 (2013)
14. Zhang, C., Smith, W.A., Dessein, A., Pears, N., Dai, H.: Functional faces: groupwise dense correspondence using functional maps. In: *CVPR*, pp. 5033–5041 (2016)