Dense Correspondence of Cone-Beam Computed Tomography Images Using Oblique Clustering Forest

Diya Sun[1]
dysun@pku.edu.cn
Yuru Pei*[1]
yrpei@pku.edu.cn
Yuke Guo[2]
guoyuke02@aliyun.com
Gengyu Ma[3]
magengyu@gmail.com
Tianmin Xu[4]
tmxuortho@gmail.com
Hongbin Zha[1]
zha@cis.pku.edu.cn

[1] Key Laboratory of Machine Perception (MOE)
Peking University
Beijing, China

[2] Luoyang Institute of Science and Technology
Luoyang, China

[3] uSens, Inc.
San Jose, USA

[4] Stomatology Hospital
Peking University
Beijing, China

Abstract

Cone-beam computed tomography (CBCT) images provide insight into underlying 3D anatomies and facilitate treatment planning and evaluations in clinical orthodontics. Establishing dense correspondence of CBCT images is a cornerstone for automatic attribute transfer and statistical shape analysis. In this paper, we propose a novel oblique clustering forest-based metric for dense correspondence of CBCT images without prior labeling. The binary partitions in branch nodes are initialized by the dominant principal component and refined by iterative point assignments. The randomly selected feature channels and the binary partition are used to learn the oblique hyperplane as the node splitting criterion. Instead of empirical forest learning, we provide a learning bound of the oblique clustering forest, which considers both the clustering compactness and the node splitting criteria determining tree traversals. We also propose a tree selection scheme under the guidance of the learning bound, which facilitates to obtain a lightweight forest. An incremental forest refinement scheme is presented to account for the growing available volumetric images. Quantitative evaluation confirms that the proposed method provides performance improvements for dense correspondences and the transfer of volumetric attributes, including landmarks and segmentation maps of CBCT images.

1 Introduction

The last decade has witnessed a rapid growth of 3D CBCT images in both the quantity and quality in clinical orthodontics to measure shape variations of craniofacial structures, such as

the mandible and teeth, due to orthodontic treatments and patient-specific growth [5, 26]. Establishing dense volumetric correspondence is a crucial step in a set of applications, including the statistical shape modeling [14] and the automatic transfer of predefined landmarks and segmentation maps [11]. The CBCT images are of relatively low signal-to-noise ratio due to limited radiation dose compared with CT images. The traditional nonrigid deformable registration methods, such as B-spline [27] and Morphons [12, 32], realize the dense volumetric correspondence. However, the volumetric deformable registration is known to be time-consuming considering the volumetric image metric and the online iterative optimizations. It is desirable to find a discriminative metric for accurate and efficient correspondence of CBCT images.

The random forest is widely used in medical imaging for its robustness and efficiency in online testing [11, 18, 28, 31, 38]. The classification and regression forests have been used to obtain correspondences of volumetric images recently [11]. In order to reduce the efforts of manual labeling, the pseudo labels obtained from the supervoxel decomposition are used to train the random forest, whereas the labels from a single volumetric image impair the generalization of the random forest. The clustering forest-based metric has been used for voxel or supervoxel-wise correspondences of CBCT images without prior labeling [22, 23]. The forest-based metrics are iteratively refined by the geodesic coordinates [22] or the mixed entropies [23]. The random forests aforementioned belong to the orthogonal forests in which one feature channel is used in the binary partition functions stored in branch nodes. The forests are composed of relatively deep decision trees with small leaves for a nice discriminative capacity, which enlarges the model complexity. On the contrary, the oblique forest [17, 34, 35] reduces the tree depth and the model complexity by virtue of oblique hyperplane-based partition functions. There already exist supervised oblique forests for tracking [35] and classification [24]. However, the underlying problem of building an oblique clustering forest without prior labeling for affinity estimation is a challenging one.

1.1 Main Contribution

In this paper, we propose an oblique clustering forest-based metric for the dense supervoxel-wise correspondence of CBCT images. We extend the splitting criterion from using traditional orthogonal hyperplanes to oblique hyperplanes to reduce the tree depth and the model complexity. The oblique clustering forest is learned from the CBCT images without prior labeling. In each node splitting, a binary partition is initialized by the signs of the dominant principal component (DPC) and refined by an iterative point assignment scheme for fast convergence. The hyperplane parameters are optimized by a ridge regression, in which the scalar parameter is optimized by maximizing the information gain of node splitting. We further give theoretical data-dependent learning bound of the oblique clustering forest considering both the clustering compactness and the complexity of leaf assignments. A tree selection scheme is proposed to obtain a lightweight forest, e.g., a forest with the number of trees below a predefined value, under the guidance of the learning bound specifically for applications with limited available memory. An incremental ridge regression method is utilized to refine the splitting criteria in branch nodes when given ever-growing available volumetric images obtained in the clinical orthodontics. It is straightforward to define a forest-based metric considering tree traversals and leaf assignments. We demonstrate the effectiveness of the proposed method in dense correspondence establishment and attribute transfer of CBCT images. In sum, the main contributions are as follows:

- An oblique clustering forest for dense volumetric correspondences of CBCT images.
- A tree selection scheme under the guidance of the learning bound of the oblique clustering forest for a lightweight model.
- An incremental forest refinement scheme to account for ever-growing available volumetric images to improve the generalization capacity.

2 Related Work

Volumetric Image Registration. Numerous works have addressed the 3D deformable image registration in the domain of medical image processing for several decades. The open-source 3D image registration packages, such as NiftyReg, are developed for head and neck registration [29]. The B-spline free-form deformation algorithm of the package realizes the block matching-based affine and deformable registrations with the normalized mutual information metric. In recent years, the demons technique using an adaptive intensity correction has been used in the registration of CBCT and CT images for the photon therapy [36]. The Morphons algorithm uses the local phase difference as a metric [12, 32] and is more robust to intensity changes than the Demons algorithm. A diffeomorphic version of the Morphons registration method has been proposed for the 3D deformable registration of CT and CBCT images [13]. However, the 3D deformable image registration using traditional volumetric image metrics, including the mutual information [15] or the normalized correlations [2], is time-consuming for head CBCT images under a large-scale optimization framework.

Random Forest-Based Correspondence. The random forest as an ensemble of decision trees is popular in the computer vision and medical image processing community for the atlas encoding [38], image segmentation [18, 31], landmark location [28], and volumetric image correspondence [11]. The regression forest-based method is utilized for the parameterizations of the anatomy localization [8]. The classification forest realizes the dense supervoxel correspondences when given pseudo-labeling of voxels from the supervoxel decomposition [11], which reduces the efforts of manual labeling. However, the generalization capacity of the random forest is impaired by the limited annotations from one volumetric image. The clustering forest is built in an unsupervised way based on the GINI impurity [4, 37] and the multivariate Gaussian distribution [7]. The clustering forest has been used for the similarity estimation of the motion trajectories [21], videos [7, 37], and volumetric images [22, 23]. Given the iteratively updated geodesic coordinates, the cascaded clustering forest is used to estimate the voxel distribution and further the voxel-wise affinity [22]. The mixed metric random forest (MMRF) introduces the weak labeling resulted from a clustering forest to discriminate the badly-clustered instances; the MMRF is fine-tuned by penalizing entropies of both the classification and the clustering random forests [23]. The aforementioned random forests using the orthogonal hyperplanes as splitting criteria are not efficient to obtain the optimal splitting planes and result in deep decision trees. Although the splitting function using one feature channel is easy to evaluate, the orthogonal decision tree often results in a blocky decision boundary instead of the desired smooth one.

Oblique Random Forest. The oblique random forest utilizes more feature channels and performs the splitting via a linear operation, which is fast to locate the optimal splitting plane and is effective to reduce the depth of the decision tree. The random forest with oblique hyperplanes is better than the orthogonal random forest in both the generalization capacity and the model compactness [17]. However, the searching space for the optimal splitting function

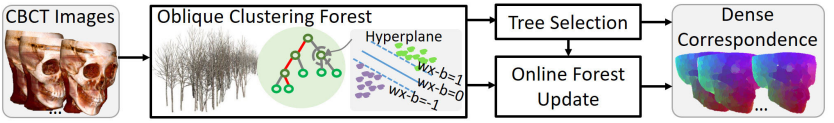

Figure 1: Flowchart of our system.

grows exponentially with hyperplane parameters. Earlier works on the oblique random forest appeared in the 90s of the last century, including CART-LC [3] and randomized search OC1 [19], which employ heuristic search methods, such as the deterministic hill-climbing, to find the oblique hyperplane. Since above methods only handle one dimension at a time, they are limited to find the optimal hyperplane. Given predefined labels, the SVM and its variants, the multi-surface proximal SVM [33], are feasible to find the optimal hyperplane, where the PCA and LDA are used to generate new feature spaces [17, 34]. The feature projection using eigenvalue decomposition in each branch node is time-consuming. An incremental oblique random forest with an online update of the hyperplane is used for object tracking [35] and image classification [24]. The oblique random forest aforementioned are mainly for the supervised classification forest instead of the unsupervised clustering forest.

3 Methods

We propose an oblique clustering forest-based method for dense correspondence of CBCT images. The input of our system is a set of craniofacial CBCT images $\mathcal{V} = \{V_{i|i=1,\ldots,N}\}$ (see Fig. 1). The goal is to estimate the pairwise dense correspondence between CBCT images. Without loss of generality, we assume each CBCT image is over-segmented into supervoxels. There is a supervoxel set $\mathcal{S} = \{s_{i|i=1,\ldots,M}\}$ related to the volumetric image dataset $\mathcal{V}$. We employ the intensity histogram to encode the supervoxel appearance. The chi-squared distances to supervoxels in the surrounding cube are also used to encode the contextual information similar to [23]. Given the supervoxel decomposition, we cast the problem as estimating supervoxel-wise distances $f : S_i \times S_j \rightarrow \mathcal{A}_{|S_i|\times|S_j|}$, where S_i and S_j are the supervoxel sets with respect to image i and j. The mapping function f in our system is defined by an oblique clustering forest composed of n_T trees for the affinity estimation, and $f = \sum_{k=1}^{n_T} f_k$. The binary metric is used, where the distance between a supervoxel pair $\mathcal{A}_k(s_i, s_j) = 0$ when they share the same traversal path to leaf nodes, and 1 otherwise.

3.1 Oblique Clustering Forest

An oblique clustering forest is an ensemble of decision trees with oblique hyperplane-based splitting criteria. Instead of the orthogonal hyperplane determined by one feature channel, the oblique decision tree utilizes a group of features channels to find the optimal hyperplane for node splitting. When given a novel instance, its traversal path is determined by the binary splitting functions stored in branch nodes. The binary splitting functions are obtained recursively in the training stage. Let $\mathcal{L}$ denote the feature family of supervoxels, including intensity-based appearance and contextual features. When building the i-th decision tree, we first randomly sample the training data $S_{i^*} \in \mathcal{S}$ with a replacement from the entire supervoxel dataset set $\mathcal{S}$. For the j-th branch node, a feature subset $L_j \in \mathcal{L}$ is randomly sampled. The data in the j-th node is denoted as $S_{i,j} \in \mathbb{R}^{|S_{i,j}|\times|L_j|}$. The binary splitting function $\tau_{i,j}$ is

defined by an oblique hyperplane, and $\tau_{i,j}(s) = [w^T s - b > 0]$ with the parameters (w, b). $[\cdot]$ is a heaviside function.

In our system, the oblique clustering forest is learned from a supervoxel set without prior labeling. Since the DPC p_* provides a continuous prior indication function of 2-way clustering [9], we initialize the binary labels as follows:

$$y(s) = \begin{cases} 1, & s^T p_* \geq 0, \\ -1, & s^T p_* < 0. \end{cases} \tag{1}$$

Since only the DPC is required, the iterative power method is utilized to solve p_* of the supervoxel set $S_{i,j}$, and $p_*^{(k)} = S_{i,j} p_*^{(k-1)} / \| S_{i,j} p_*^{(k-1)} \|$ in the $k-$th iteration. Instead of using Forgy's random partition [10] for centroid seeds, we initialize seeds by the DPC, $\mu_{k|k=\pm 1} = 1/n_k \sum_{y(s)=k} s$. n_k denotes the number of supervoxels in cluster k. We perform the centroid update and point assignments iteratively to refine the 2-way clustering as the k-means method. The sign function (Eq. 1) defined by the DPC helps to realize fast convergence of the binary partition in our experiments.

The linear SVM as a binary classifier is suitable for node splitting. The normal vector $w \in \mathbb{R}^{|L_j|}$ and a scalar $b \in \mathbb{R}$ determine the hyperplane. The samples are split into the left or right children nodes according to which side of the hyperplane they lie. Given the previous labeling, the hyperplane parameters are learnt by minimizing $\frac{\lambda}{2}(\|w\|^2 + b^2) + \frac{1}{2}\sum_{i=1}^{|S_{i,j}|} \max(0, 1 - y_i(w^T s_i - b))$, where λ is a regularization parameter. The ridge regression gives a close-form solution of the above problem.

$$w_* = (S_{i,j}^T S_{i,j} + \lambda I)^{-1} S_{i,j}^T y, \tag{2}$$

where $w_* = [b\ w]$. In order to maximize the information gain of the node splitting, the hyperplane is refined by exhaustively searching the scalar b to maximize the information gain I. $I_{i,j} = -\sum_{k=0,1} \alpha_k tr(\Sigma(S_{i,j}^k))$. The trace operator is used to tackle the ubiquitous rank deficiency of the covariance matrix Σ of data $S_{i,j}^k$ in left and right children nodes [21]. The scalar weight $\alpha_k = |S_{i,j}^k| / \sum_{k=0,1} |S_{i,j}^k|$, which is defined by the node cardinality. We randomly sample a set of scalar values of b, and identify the one to maximize I. The procedure to build the oblique clustering forest is summarized in Algorithm 1.

Algorithm 1 Oblique Clustering Forest

Input:Supervoxel set $\mathcal{S}$, forest size n_T
Output:Oblique clustering forest

```
for i = 1 : n_T do
    Randomly sample a supervoxel set S_{i*} ∈ S ;
    for Each node splitting do
        Estimate the DPC and initialize the binary partition (Eq. 1);
        Refine the binary partition;
        Estimate the hyperplane (Eq. 2);
        Exhaustively search the scalar b and identify the one maximizing the information gain;
        Split the current node;
    end for
end for
```

3.2 Generalizability and Tree Selection

Despite the empirical success of the forest-based metric, a major challenge is to get theoretical guarantees about the efficiency and quality of the forest-based metric for the dense correspondence. We analyze the complexity of the clustering hypothesis of the oblique clustering forest and give a data-dependent learning bound. Leaf nodes of each oblique decision tree provide a partition of the data. The loss function of the clustering compactness with n_l leaf node $h(s) = \sum_{j=1}^{n_l} r(s,l_j)u_j(s)$, and $u_j(s) = 1/n_j\|s-\mu_j\|^2$. n_j and μ_j denote the number of instances and the centroid of the leaf j. The binary function r indicates the leaf assignment, and $r(s,l) = 1$ when supervoxel s is assigned to the leaf node l. The hypothesis of the oblique clustering tree is defined by the centroid vector $(\mu_1,\ldots,\mu_{n_l})$ of n_l leaf nodes. Let $\mathcal{D}$ denote the data domain. The training supervoxel data $\mathcal{S} \in \mathcal{D}$.

Definition 3.1. For an independent distributed data set $S = \{s_{i|i=1,\ldots,n}\}$, and $S \in \mathcal{D}$, if $\mathcal{H}$ is a class of clustering hypotheses of oblique trees on $\mathcal{D}$, the Rademacher complexity

$$\mathfrak{R}(\mathcal{H}) = E_\sigma \sup_{h\in\mathcal{H}} \sum_{i=1}^{n} \frac{1}{n}\sigma_i h(s_i) = E_\sigma \sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i \sum_{j=1}^{n_l} r(s_i,l_j)\|s_i-\mu_j\|^2/n_j. \tag{3}$$

where σ is an i.i.d random variable set to 1 or -1 with the probability $p \in [0,\frac{1}{2}]$, an 0 with the probability $1-2p$.

Theorem 3.2. *For an oblique clustering tree, any independently distributed instance set* $S = \{s_{i|i=1,\ldots,n}\}$, $S \in \mathcal{D}$, *and* $\delta > 0$, *with probability of at least* $1-\delta$, *the following holds*

$$E_x h(x) \le \frac{1}{n}\sum_{i=1}^{n} h(s_i) + \sum_{j=1}^{n_l}\left(\underbrace{\frac{4\sqrt{\pi/2}a_j^2}{n_j^{\frac{3}{2}}}}_{\textit{Leaf Compactness}} + \underbrace{\sqrt{\frac{2\nu_j(\beta\log\frac{e}{\eta}+\log(\sqrt{\beta}+Ba_j)+\log\frac{2}{\sqrt{n}})}{n}}}_{\textit{Leaf Assignment}}\right). \tag{4}$$

The learning bound (Eq. 4) of the oblique clustering tree includes two terms regarding the clustering compactness and the leaf assignments. a_j denotes the data radius of the leaf node l_j, and $\|s\| \le a_j$ when $r(s,l_j) = 1$. ν_j denotes the depth of the leaf j. β is the cardinality of feature channels used in estimating splitting criteria and set to $|L|$. $\eta = \beta/|\mathcal{L}|$ regarding a feature family $\mathcal{L}$. The first term on the leaf compactness is determined by the leaf radius a_j and the leaf size n_j. The compact leaf nodes imply that the data-dependent upper bound of the oblique clustering tree is low. The second term is associated with the leaf assignment of oblique clustering trees, which is related to the principal-component-based binary partition and the linear SVM-based hyperplane estimation. The β-dimensional principal component analysis has a Rademacher complexity of $2\sqrt{\beta/n}$ [16]. The linear SVM has a complexity of $2Ba_j/\sqrt{n}$, and $\|w\| \le B$ [25]. There are $\binom{|\mathcal{L}|}{\beta}$ ways to choose β feature channels from $\mathcal{L}$, and $\binom{|\mathcal{L}|}{\beta} \le \left(\frac{|\mathcal{L}|e}{\beta}\right)^\beta$. According to Massart's lemma, the Rademacher complexity of the leaf assignments $\mathfrak{R}(\mathcal{R}_j) \le \sqrt{(2\sum_{k=1}^{\nu_j}\log\Lambda r_k)/n}$ in terms of the growth function Λr_k.

Since $\log \Lambda r_k \leq \log \binom{|\mathcal{L}|}{\beta} \left(2\sqrt{\beta/n} + 2Ba_j/\sqrt{n}\right)$, we derive the second term of the learning bound regarding the leaf assignment as in Eq. 4. Since oblique clustering trees tend to be shallow and have large leaves compared with the orthogonal decision trees, the learning bound implies that the oblique tree has lower complexity than the orthogonal one.

Tree Selection. Given the learning bound in Theorem 3.2, we propose a tree selection scheme to generate a lightweight forest, e.g., a forest with the number of trees below a predefined value, which helps to downsize the forest model specifically for the applications with limited available memory. Considering learning a clustering forest with n_T trees, we first construct the oblique forest with $\tilde{n}$ trees $(\tilde{n} > n_T)$ as described in algorithm 1. Second, the learning bounds (Eq. 4) of all oblique trees are sorted. Finally, the first n_T trees with the lowest learning bound are selected to build the final oblique clustering forest. As Breimann has stated in his pioneering work, the forest surely converges with increasing forest size [4]. However, when given limited memory in online applications, we would like to reduce the forest size. The learning bound as stated in Theorem 3.2 gives a guide for the tree selection. The oblique clustering tree with the small learning bound is prone to be shallow and with compact leaf nodes. The tree selection brings additional computational complexities regarding the learning bound estimation. Since the learning bound only relies on the leaf size n_j, the data radius a_j, the tree depth v_j, and the number of feature channels β used in node splitting, the learning bound estimation is efficient in our system.

3.3 Online Forest Update

It is intuitive that the online forest update given the newly available volumetric images is helpful to improve the generalization capacity. We propose an incremental update method of the oblique clustering forest. When given new CBCT images $\hat{V}$ and accompanying supervoxel set $\hat{S}$, the node splitting functions are updated accordingly. The model update starts from the root node. We first incrementally update the DPC using Oja's algorithm [20]. When given new data $\hat{S}_{i^*}$ of the tree i, the newly updated DPC $p_*^{(t+1)}$ of the root node is obtained as follows:

$$p_*^{(t+1)} = \frac{p_*^{(t)} + \gamma \hat{S}_{i^*}^T \hat{S}_{i^*} p_*^{(t)}}{\| p_*^{(t)} + \gamma \hat{S}_{i^*}^T \hat{S}_{i^*} p_*^{(t)} \|}, \tag{5}$$

where γ is a learning rate set at $\frac{1}{|S_{i^*}|+|\hat{S}_{i^*}|}$. The instance labels of the newly added data are defined as

$$\hat{y}(\hat{s}) = \begin{cases} 1, \ \hat{s}^T p_*^{(t+1)} \geq 0 \text{ and } \arg\min_k \|\hat{s} - \mu_k\| = 1, \\ -1, \ \hat{s}^T p_*^{(t+1)} < 0 \text{ and } \arg\min_k \|\hat{s} - \mu_k\| = -1. \end{cases} \tag{6}$$

$\mu_{k|k=\pm 1}$ denotes the clustering centroid, which is updated when given the additional data. $\mu_k^{(t+1)} = \frac{|S_{i^*}^k|}{|S_{i^*}^k|+|\hat{S}_{i^*}^k|}\mu_k^{(t)} + \frac{|\hat{S}_{i^*}^k|}{|S_{i^*}^k|+|\hat{S}_{i^*}^k|}\hat{\mu}_k$, and $\hat{\mu}_k = \frac{1}{|\hat{S}_{i^*}^k|}\sum_{\hat{y}(\hat{s})=k} \hat{s}$. $|S_{i^*}^k|$ and $|\hat{S}_{i^*}^k|$ are the cardinalities of the supervoxel set $S_{i^*}^k$ used to estimate the hyper-parameters w_* and the newly added data $\hat{S}_{i^*}^k$ of cluster k. The hyper-parameters are updated when given the new data $(\hat{S}_{i^*}, \hat{y})$.

$$w_* = \left(\begin{bmatrix} S_{i^*} \\ \hat{S}_{i^*} \end{bmatrix}^T \begin{bmatrix} S_{i^*} \\ \hat{S}_{i^*} \end{bmatrix} + \lambda I \right)^{-1} \begin{bmatrix} S_{i^*} \\ \hat{S}_{i^*} \end{bmatrix}^T \begin{bmatrix} y \\ \hat{y} \end{bmatrix}. \tag{7}$$

The incremental update of hyper-parameters can be rewritten as

$$w_*^{(t+1)} = w_*^{(t)} + C^{(t)} w_*^{(t)} + A^{(t+1)} \hat{S}_{i^*}^T \hat{y}, \tag{8}$$

where $A^{(0)} = (S_{i^*}^T S_{i^*} + \lambda I)^{-1}$, and $A^{(t+1)} = A^{(t)} - A^{(t)}\hat{S}_{i^*}^T(I + \hat{S}_{i^*}A^{(t)}\hat{S}_{i^*}^T)^{-1}\hat{S}_{i^*}A^{(t)}$. $C^{(t)} = -A^{(t)}\hat{S}_{i^*}^T(I + \hat{S}_{i^*}A^{(t)}\hat{S}_{i^*}^T)^{-1}\hat{S}_{i^*}$. The formulation (Eq. 8) only make use of the newly added data $(\hat{S}_{i^*}, \hat{y})$ to update the parameters of the oblique hyperplane. We only need to store $A^{(t)}$, the cardinality, and the centroid $\mu_{\pm 1}$ of the binary partition. Given the updated splitting criteria in the root node, S_{i^*} is recursively assigned to left and right children nodes. The newly added data are assigned binary labels (Eq. 6), and then used to update the hyperplanes (Eq. 7 and 8).

4 Experiments

The dataset includes 50 clinically captured CBCT images. The CBCT images are captured by NewTom scanner and re-sampled to a resolution of $250 \times 250 \times 238$. The voxel size is $0.8 \times 0.8 \times 0.8\, mm^3$. CBCT images are decomposed to supervoxels using the SLIC algorithm [1]. Each CBCT image is decomposed into $20k$ supervoxel; the compactness parameter is set at 10. We employ the proposed oblique clustering forest to estimate pairwise affinities and correspondences between CBCT images, where the supervoxel pair sharing the same traversal path are assumed to be similar.

Segmentation Map Transfer. We perform the segmentation map transfer of the mandible and the maxilla. We evaluate the map transfer by the Dice similarity coefficient (DSC) and the average Hausdorff distance (AHD) as shown in Table 1. We compare with the binary partition and hyperplane learning using the DPC only (Eq. 1), with iterative refinement (IR), with additional scalar searching (IR-SS), the oblique forest with the tree selection (OFTS) (Section 3.2), and the oblique forest with the online update (OFOU) (Section 3.3). As we can see, the iterative refinement of the binary partition and the scalar searching are helpful to improve the transfer accuracies. For the forest with 50 trees, the OFTS outperforms the IR-SS without tree selections. We think the reason is that the trees minimizing the leaning bound tend to have compact leaves and help to identify the accurate correspondences. We also compare with the OFOU, where the splitting criteria are updated when given newly available CBCT images. Note the tree structures do not change during the online update. The newly available images are helpful to improve the correspondence solving as shown in Table 1. Furthermore, the proposed method improves the separation of the mandible and the maxilla when faced with the intercuspation as shown in Fig.2 (with error regions blocked).

Table 1: Segmentation map transfer using the DPC, IR, IR-SS, OFTS, and OFOU methods.

	DSC (%)					AHD ($\times 10^{-1}$ mm)				
	DPC	IR	IR-SS	OFTS	OFOU	DPC	IR	IR-SS	OFTS	OFOU
Mandible	87.9	89.1	90.6	92.0	**92.9**	4.45	4.32	4.03	3.92	**3.30**
Maxilla	93.0	94.1	94.5	95.2	**95.4**	3.14	2.89	2.45	2.44	**2.41**

Landmark Transfer. Given the dense correspondences between images, it is intuitive to transfer the predefined landmarks from the reference image to novel ones. The transfer accuracy is given by $e = n_q/n_a$. n_a denotes the total number of testing images. n_q denotes the number of testing images in which the supervoxel enclosing the ground truth landmark occurs in the q nearest matching supervoxels. The transfer results of four landmarks, i.e. *ec, zy, ns, gn*, are shown in Fig. 3 (a-d). Similar to the segmentation map transfer, the IR and the SS are helpful to improve the oblique hyperplane learning. Moreover, given a fixed forest size, the oblique forest obtained using tree selections outperforms the one without tree se-

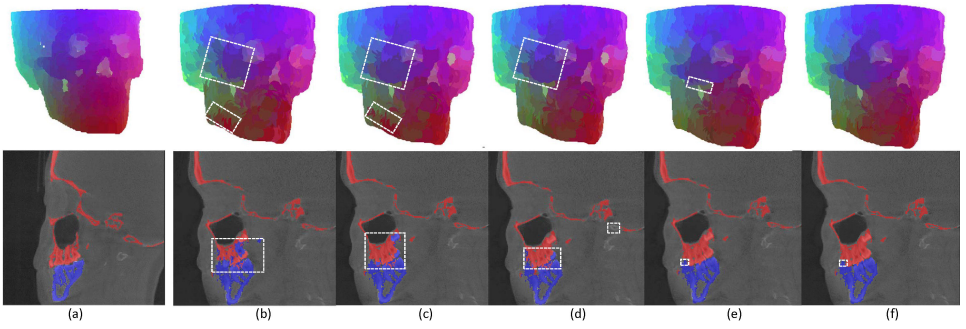

Figure 2: Correspondences (top) and segmentation map transfer (bottom) from (a) the reference to the target image using (b) DPC, (c) IR, (d) IR-SS, (e) OFTS, and (f) OFOU methods.

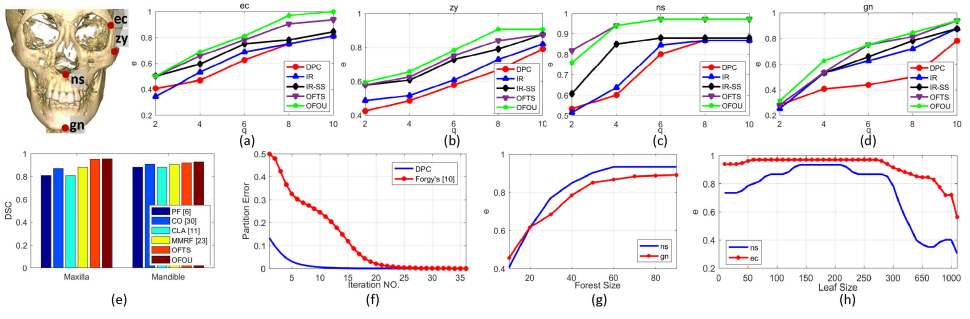

Figure 3: Landmark transfer accuracy of (a) ec, (b) zy, (c) ns, and (d) gn with q ranging from 2 to 10. (e) Comparisons with the CO [30], the PF [6], the CLA [11], and the MMRF [23] methods. (f) The binary partition errors with increasing numbers of iterations. Landmark transfer accuracies using (g) different numbers of trees and (h) sizes of leaf nodes.

lections. The OFOU facilitates to improve the generalization capacity and correspondences, achieving the best performances in almost all landmark transfer experiments.

Compare with Other Methods. The proposed method is compared with the state-of-the-art label transfer methods, including the convex optimization (CO) [30], the patch-based fusion (PF) [6], the classification forests (CLA) [11], and the MMRF [23] (Fig.3(e)). The proposed method achieves comparable performances in the segmentation map transfer of the maxilla and the mandible. Note the oblique forest is more shallow than other forest-based methods. The average leaf size of the oblique clustering forest is 200 vs. 5-20 of the orthogonal clustering forests [23], which helps to reduce the model complexity.

Parameter Analysis. We use the iterative point assignment to refine the binary partition initialized by the DPC, which facilitate a fast convergence compared with the Forgy's random partition (Fig. 3(f)). As Breimann stated in his pioneering work, the random forest almost surely converges when given an increasing number of trees [4]. As shown in Fig. 3(g), the landmark transfer accuracies increase with the forest size. The oblique forest is known to be more feasible to get the optimal binary node splitting criteria. The oblique forest produces good transfer performances with a comparatively large leaf size, set to 200 in our experiments, and a low model complexity (see Fig. 3(h)).

5 Conclusions

In this paper, we propose a novel correspondence solving approach of CBCT images using the oblique clustering forest. By virtue of the tree selection under the guidance of a learning bound of oblique decision trees and an online update scheme, the proposed method is feasible to establish correspondences between volumetric images and further attribute transfer. Experiments demonstrate the proposed method is effective in both reducing model complexities and achieving performance improvements compared with state-of-the-arts.

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

References

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.

[2] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.

[3] L. I. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. Classification and regression trees (cart). *Encyclopedia of Ecology*, 40(3):582–588, 1984.

[4] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[5] Lucia H. S. Cevidanes, Martin A. Styner, and William R. Proffit. Image analysis and superimposition of 3-dimensional cone-beam computed tomography models. *American Journal of Orthodontics and Dentofacial Orthopedics*, 129(5):611, 2006.

[6] Pierrick Coupé, José V Manjón, Vladimir Fonov, Jens Pruessner, Montserrat Robles, and D Louis Collins. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2):940–954, 2011.

[7] A Criminisi, J Shotton, and E Konukoglu. Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. *Microsoft Research Cambridge, Tech. Rep. MSRTR-2011-114*, 5(6):12, 2011.

[8] Antonio Criminisi, Jamie Shotton, Duncan Robertson, and Ender Konukoglu. Regression forests for efficient anatomy detection and localization in ct studies. In *International MICCAI Workshop on Medical Computer Vision*, pages 106–117. Springer, 2010.

[9] Chris Ding and Xiaofeng He. K -means clustering via principal component analysis. *Proc. Int. Conf. Mach. Learning, 2004*, 46(4):29, 2004.

[10] Greg Hamerly and Charles Elkan. Alternatives to the k-means algorithm that find better clusterings. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 600–607. ACM, 2002.

[11] Fahdi Kanavati, Tong Tong, Kazunari Misawa, Michitaka Fujiwara, Kensaku Mori, Daniel Rueckert, and Ben Glocker. Supervoxel classification forests for estimating pairwise image correspondences. *Pattern Recognition*, 63:561–569, 2017.

[12] Hans Knutsson and Mats Andersson. Morphons: Paint on priors and elastic canvas for segmentation and registration. *Image Analysis*, pages 46–55, 2005.

[13] Guillaume Landry, Reinoud Nijhuis, George Dedes, Josefine Handrack, Christian Thieke, Guillaume Janssens, Jonathan Orban de Xivry, Michael Reiner, Florian Kamp, Jan J Wilkens, et al. Investigating ct to cbct image registration for head and neck proton therapy as a tool for daily dose recalculation. *Medical physics*, 42(3):1354–1366, 2015.

[14] Herve Lombaert, Michael Arcaro, and Nicholas Ayache. *Brain Transfer: Spectral Analysis of Cortical Surfaces and Functional Maps*. Springer International Publishing, 2015.

[15] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

[16] Andreas Maurer and Massimiliano Pontil. k -dimensional coding schemes in hilbert spaces. *IEEE Transactions on Information Theory*, 56(11):5839–5846, 2010.

[17] Bjoern H. Menze, B. Michael Kelm, Daniel N. Splitthoff, Ullrich Koethe, and Fred A. Hamprecht. On oblique random forests. In *European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 453–469, 2011.

[18] Albert Montillo, Jamie Shotton, John Winn, Juan Eugenio Iglesias, Dimitri Metaxas, and Antonio Criminisi. Entangled decision forests and their application for semantic segmentation of ct images. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 184–196. Springer, 2011.

[19] Sreerama K. Murthy, Simon Kasif, Steven Salzberg, and Richard Beigel. Oc1: A randomized algorithm for building oblique decision trees. 1993.

[20] E. Oja. Subspace methods of pattern recognition. *Signal Processing*, 7(1):79, 1983.

[21] Yuru Pei, Tae-Kyun Kim, and Hongbin Zha. Unsupervised random forest manifold alignment for lipreading. In *IEEE International Conference on Computer Vision*, pages 129–136, 2013.

[22] Yuru Pei, Yunai Yi, Gui Chen, Tianmin Xu, Hongbin Zha, and Gengyu Ma. Voxel-wise correspondence of cone-beam computed tomography images by cascaded randomized forest. In *IEEE International Symposium on Biomedical Imaging*, pages 481–484, 2017.

[23] Yuru Pei, Yunai Yi, Gengyu Ma, Yuke Guo, Gui Chen, Tianmin Xu, and Hongbin Zha. Mixed metric random forest for dense correspondence of cone-beam computed tomography images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017.

[24] M Ristin, M Guillaumin, J Gall, and Gool L Van. Incremental learning of random forests for large-scale image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):490–503, 2016.

[25] John Shawe-Taylor and Nello Cristianini. Kernel methods for pattern analysis. *Publications of the American Statistical Association*, 101(476):1730–1730, 2004.

[26] M. A. Silva, U Wolf, F Heinicke, A Bumann, H Visser, and E Hirsch. Cone-beam computed tomography for routine orthodontic treatment planning: a radiation dose evaluation. *American journal of orthodontics and dentofacial orthopedics : official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics*, 133(5):640.e1, 2008.

[27] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153, 2013.

[28] Neslisah Torosdagli, Denise K Liberton, Payal Verma, Murat Sincan, Janice Lee, Sumanta Pattanaik, and Ulas Bagci. Robust and fully automated segmentation of mandible from ct scans. In *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 1209–1212, 2017.

[29] Catarina Veiga, Jamie McClelland, Syed Moinuddin, Ana Lourenço, Kate Ricketts, James Annkah, Marc Modat, Sébastien Ourselin, Derek DąśSouza, and Gary Royle. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using ct-to-cbct deformable registration for ąřdose of the dayąś calculations. *Medical physics*, 41(3), 2014.

[30] Li Wang, Ken Chung Chen, Feng Shi, Shu Liao, Gang Li, Yaozong Gao, Steve GF Shen, Jin Yan, Philip KM Lee, Ben Chow, et al. Automated segmentation of cbct image using spiral ct atlases and convex optimization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 251–258. Springer, 2013.

[31] Li Wang, Yaozong Gao, Feng Shi, Gang Li, Ken-Chung Chen, Zhen Tang, James J Xia, and Dinggang Shen. Automated segmentation of dental cbct image with prior-guided sequential random forests. *Medical physics*, 43(1):336–346, 2016.

[32] Andreas Wrangsjö, Johanna Pettersson, and Hans Knutsson. Non-rigid registration using morphons. *Image Analysis*, pages 501–510, 2005.

[33] Le Zhang and Ponnuthurai N. Suganthan. Oblique decision tree ensemble via multisurface proximal support vector machine. *IEEE Transactions on Cybernetics*, 45(10): 2165, 2015.

[34] Le Zhang and Ponnuthurai Nagaratnam Suganthan. Random forests with ensemble of feature spaces. *Pattern Recognition*, 47(10):3429–3437, 2014.

[35] Le Zhang, Jagannadan Varadarajan, Ponnuthurai Nagaratnam Suganthan, Narendra Ahuja, and Pierre Moulin. Robust visual tracking using oblique random forests. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5825–5834, 2017.

[36] Xin Zhen, Xuejun Gu, Hao Yan, Linghong Zhou, Xun Jia, and Steve B Jiang. Ct to cone-beam ct deformable registration with simultaneous intensity correction. *Physics in medicine and biology*, 57(21):6807, 2012.

[37] Xiatian Zhu, Chen Loy, and Shaogang Gong. Video synopsis by heterogeneous multi-source correlation. In *IEEE International Conference on Computer Vision*, pages 81–88, 2013.

[38] Darko Zikic, Ben Glocker, and Antonio Criminisi. Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. *Medical Image Analysis*, 18(8):1262–1273, 2014.