Stratified Decision Forests for Accurate Anatomical Landmark Localization in Cardiac Images

Ozan Oktay*, Wenjia Bai, Ricardo Guerrero, Martin Rajchl, Antonio de Marvao, Declan P. O'Regan, Stuart A. Cook, Mattias P. Heinrich, Ben Glocker, and Daniel Rueckert

Abstract—**Accurate localization of anatomical landmarks is an important step in medical imaging, as it provides useful prior information for subsequent image analysis and acquisition methods. It is particularly useful for initialization of automatic image analysis tools (e.g. segmentation and registration) and detection of scan planes for automated image acquisition. Landmark localization has been commonly performed using learning based approaches, such as classifier and/or regressor models. However, trained models may not generalize well in heterogeneous datasets when the images contain large differences due to size, pose and shape variations of organs. To learn more data-adaptive and patient specific models, we propose a novel stratification based training model, and demonstrate its use in a decision forest. The proposed approach does not require any additional training information compared to the standard model training procedure and can be easily integrated into any decision tree framework. The proposed method is evaluated on 1080 3D high-resolution and 90 multi-stack 2D cardiac cine MR images. The experiments show that the proposed method achieves state-of-the-art landmark localization accuracy and outperforms standard regression and classification based approaches. Additionally, the proposed method is used in a multi-atlas segmentation to create a fully automatic segmentation pipeline, and the results show that it achieves state-of-the-art segmentation accuracy.**

Index Terms—**Automatic landmark localization, cardiac image analysis, multi-atlas image segmentation, stratified forests.**

I. Introduction

ACCURATE detection of anatomical landmarks is important for clinical applications that require fully-automated

image segmentation and registration. In many cases, landmark localization is a prerequisite for the initialization of subsequent image analysis steps, such as initialization of deformable model and atlas based approaches [1] for cardiac modelling. Similarly, detected landmarks can be used to facilitate fully automatic planning of image acquisitions, such as cardiac MRI examinations [2]. Additionally, landmark localization in cardiac images, e.g. right ventricle (RV) insertion points, can be used to analyse left ventricle function according to AHA myocardial segmentation models [3]. However, variable imaging quality and variations of the heart's shape, size and orientation across subjects and populations pose a great challenge to fully automatically detect landmarks from medical images, particularly in cardiac MR images.

A common approach to locate landmarks of interest is based on predictions of a trained classifier or regressor. This achieves reliable and accurate results as long as both the learned model and the training data are similar enough. However, existing approaches mostly use averages or other simple statistics over subsets of training examples which can yield large prediction errors on test data. In this paper, we propose a generic learning approach (stratification) that can be used in the context of decision trees to learn more adaptive classifiers and regressors without requiring additional training information. With the proposed method, models are trained not only with local information collected from image patches, but also with global information such as shape, size, and pose differences of the organs. Indeed, it is the main motivation for the use of the term stratification for the proposed method. In this way, the landmark detection accuracy can be improved as shown in Figure 1: This example shows that the proposed method can localize the anatomical landmarks accurately although the organ of interest (in this case the heart) exhibits large variations in terms of pose, size and shape.

II. Related Work

Landmark localization has been a particular research area in both computer vision and medical image analysis over the last decades. Landmark localization methods in both medical and natural images can be mainly categorized into two subgroups, namely image registration and model based approaches. The approaches in the former category identify dense correspondences between target and atlas images to localize anatomical landmarks. Robust image alignment [4]–[6] and non-rigid
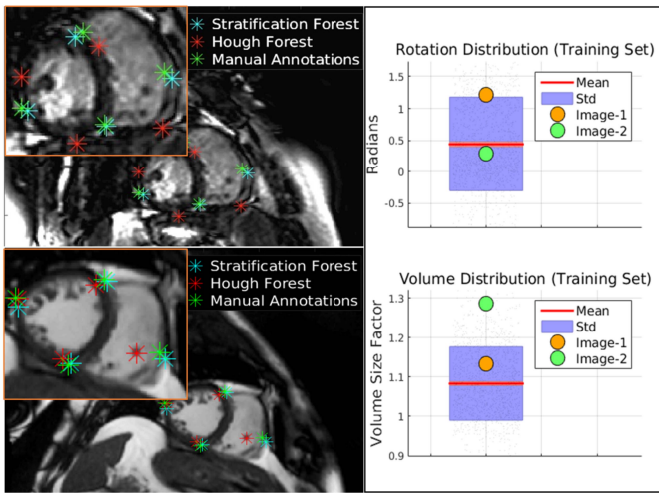
Fig. 1. A comparison between standard Hough and proposed stratified decision forests for landmark localization in two different short axis cardiac MR images. The first image (top) has a significantly different heart orientation than the population average shown by the red bar (right); similarly the heart in the second image (bottom) is larger in terms of size in comparison to the average heart size in the training images. The proposed approach is able to cope with these variations by learning adaptive and patient-specific regression models.

registration [7], [8] techniques can be used to relate anatomical correspondences between images.

The approaches in the later category, on the other hand, learn discriminative and generative models through training datasets to associate anatomical landmark locations with prior information, such as shape and appearance information of tissues. These approaches can be further categorised into three subgroups as classification, regression, and graphical models [9]. In addition to these categories, there are several examples such as Hough Forests [10] in which the regression and classification approaches are combined in a hybrid manner. In the first category, work based on the marginal space learning (MSL) framework [11] proposed the use of cascaded boosting tree detectors to identify pose and landmark locations in medical images. The presented cascaded classifiers method [2], [12] is applied on 2D cardiac MR images to detect landmarks. In computer vision, Shotton et al. [13] used trained classifier trees to identify body parts in depth images, which is used as an intermediate step to locate joint positions in human body. Similarly, Donner et al. [14] proposed a decision forest based classifier to locate hand joint locations in X-ray images. More recent work [15] proposed the use of separable filters that are learned via multi-layer perceptrons, in boosting tree classifiers. However, the main limitation of classifier based approaches is caused by the limited number of positive training samples. Moreover, classifier based approaches are susceptible to imaging artefacts and low imaging quality in image regions where the landmarks are located. In ultrasound images, shadowing [16] can be another limitation for the classifier based approaches, which is similar to occlusions in natural images.

To avoid these problems, the regression based methods combine the knowledge from different image regions that can be near or distant from the target landmark locations. Aggregation of predictions from several regions can yield more robust

localization and addresses the problems mentioned above. Particularly, for the skeletal joint localization problem, recent methods [17], [18] proposed regression based localization approaches and showed an improved performance over the classification methods. Similarly, in the context of medical image analysis, Criminisi et al. [19] have used regression trees to locate 3D bounding boxes corresponding to organs in abdominal CT images. In a recent work [20], the regression formulation is modified by changing the standard offset model parameters with atlas scale and position regression parameters. This was shown to achieve improved localization performance in CT images as it makes use of global shape information of organs.

Hybrid approaches, on the other hand, integrate regression and classification models in a single detector framework. These approaches benefit from richer annotations as the regression models are conditioned on segmentation labels. In [21], a joint model is proposed to locate vehicles in natural images. Similarly, Hough forests [10] are used for pedestrian and horse detection in natural images.

The common parametric model used in regression models is based on multivariate normal distributions. For this reason, the learned models may not be specific enough for all cases, in particular where the sub-populations form multiple clusters, e.g. different heart size and shape in medical images. Similarly, in the context of facial image analysis, the pose of the human head is an important latent variable that can affect the model performance. For facial gesture recognition, the methods presented in [22], [23] address this problem by proposing conditional and hierarchical decision trees. However, the conditional trees require additional classifiers to estimate the probability of this latent variable, and based on its value the required decision trees are selected for inference. Similarly, spectral forest methods [24] have been proposed to allow population specific bagging to train specialized decision trees, which showed an improved segmentation performance compared to the standard bagging. However, the proposed tree selection process in testing requires the computation of nearest neighbours for each test image as a pre-processing step.

In this paper, we present a single unified decision tree training approach that generalizes the previously presented cascaded localization frameworks. Particularly, the latent variable parameters are computed within the stratified trees in a probabilistic way rather than using auto-context models as in MSL. Additionally, the proposed approach does not require any additional training information, specialized decision trees [22], or externally applied dimensionality reduction methods [24] for training and inference. Our method is along the same line as the previously proposed neighbourhood approximation forests [25], where anatomically similar training images are grouped together in the sub-tree architecture to obtain population clusters. Differently, in the proposed method, we show that this type of clustering with stratification nodes can enhance the classification and regression performance without adding significant computational costs. The proposed novelties in this paper can be characterized as follows:

- A stratification training objective is presented to learn more data adaptive decision tree models in an implicit

TABLE I
A SHORT SUMMARY OF RELATED WORK, DECISION FOREST BASED
CLASSIFICATION AND REGRESSION APPROACHES IN THE LITERATURE

Related Work	Method Description
Criminisi et al. [10]	A regression forest based landmark detection method is applied to locate visceral organs in full-body CT images.
Gall et al. [18]	Hough forest method is presented for joint training of classification and regression nodes.
Konukoglu et al. [23]	Neighbourhood approximation forest (NAF) is proposed to cluster brain MR images based on subjects' age and pathology group. In testing time, the learned models are used to find most similar subjects among the training images.

way. In this way, the learned model can be more representative for datasets showing large variations such as object shape and size.

- The proposed model does not require cascaded classifiers [12] or specialized trees [22], [24] to learn the statistics of latent variables, such as pose parameters. Therefore, it can be considered as a generalization of cascaded models and can be more easily adapted to other classification/regression problems.
- The presented work shows the application of joint classification and regression based approach on anatomical landmark localization in cardiac MR images.

Besides the proposed novelties, the method presented in this paper is built on some existing classification and regression techniques in the literature (see Table I). The improvements upon these existing approaches can be described as follows:

- Criminisi et al. [19]: Regression splits are defined in a similar way, but the leaf node prediction models are characterised by the prior information which is the probability of class labels and global shape characteristics.
- Gall et al. [10]: In addition to classification and regression labels, global characteristics of organs (e.g. shape/size/pose) are incorporated in a tree model to learn more representative regression models.
- Konukoglu et al. [25]: We show that clustering of training samples based on their global characteristics, such as size and shape, can actually improve classification and regression accuracy. In contrast to this, [25] is used in subject clustering and nearest-subject search problems.

The proposed method is validated on 1080 3D and 90 multi-stack 2D short-axis cardiac MR images acquired with different scanners. The results show that it achieves more accurate localization results compared to Hough forests and MSL based cascaded classifier localization methods. The experiments also show that it provides better initialization for subsequent multi atlas image segmentation [26].

III. THEORY

A decision tree is a tree-structured predictor consisting of two types of nodes, split nodes $\psi \in \mathcal{T}$ and leaf nodes $\ell \in \mathcal{L}$. Split nodes route samples $x \in \mathcal{X}$ to leaf nodes to find the best match for a given sample against a set of training examples, whereas leaf nodes store the posterior distributions $p(y|x)$ for output variable $y \in \mathcal{Y}$ and make a prediction for the sample $x \in \mathcal{X}$. Split nodes are characterized by decision stumps $\psi(x) : \mathcal{X} \rightarrow \{0, 1\}$ which route samples to left and right sub-decision trees. An ensemble of uncorrelated decision tree classifiers is known as a decision forest [27].

In this paper, a new stratification training objective is proposed to learn node split models that can group samples based on their global characteristics such as organ shape, size, and orientation. The presented training scheme includes structured classification, regression, and stratification split nodes, which will be explained in the following sections.

A. Input Space

The input space of the decision trees is characterized by image channels $\mathcal{I}$, and there are C channels for each training sample $\mathcal{I}_i = (I_i^1, I_i^2, \cdots, I_i^C)$. The channels are defined by multi-resolution appearance, histograms of gradients (HoG) [28], and gradient magnitude image patches of size $(M_a)^3$. The appearance channels are formed by constructing a two-scale Gaussian image pyramid, where the original input is downsampled by factor of 2 in the top layer. The smoothed gradient magnitude and orientation are computed with oriented Gaussian derivative kernels. Orientation histograms are computed using soft-binning, where bin weights are determined by the gradient magnitude (cf. [29]). The features for each patch $\mathcal{P}_i$ centred at $\boldsymbol{p}_c$ are extracted from each channel $\alpha \in \{1, 2, \cdots, C\}$ by performing comparisons of the average intensity of boxes (R_1, R_2) within the patch boundaries in a similar fashion to [13], [30].

$$q_\lambda(\boldsymbol{p}_c) = \frac{1}{|R_1|} \sum_{\boldsymbol{p} \in R_1} I^\alpha(\boldsymbol{p}) - \frac{1}{|R_2|} \sum_{\boldsymbol{p} \in R_2} I^\alpha(\boldsymbol{p}) \quad (1)$$

For a single dimensional decision stump, the split node parameters are defined by the parameter set $\lambda = (R_1, R_2, \alpha)$ and threshold value γ. For pairwise channel comparisons, the box regions R_i are chosen to be non-zero, and for single channel comparisons R_2 is chosen to be zero. Using the same notation, the split function is defined as:

$$\psi(q_\lambda(\boldsymbol{p}_c)) = \begin{cases} 1, & \text{if } q_\lambda(.) \geq \gamma \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

B. Structured Classification

In the proposed method, the information stored in the leaf nodes (output space $\mathcal{Y}$) is characterized by class labels, regression, and stratification model parameters. In other words, a single decision tree model is learned to perform multiple tasks simultaneously, such as organ surface delineation, landmark location regression, and shape information regression. The joint training of classification and regression nodes benefit from each other to learn more representative and class specific models since more ground-truth information is provided during training as suggested by Gall et al. [10].

For the organ surface delineation task, the class labels are stored in the leaf nodes as label patches by applying the

structured learning methods proposed in [31], [32] which have been shown to increase the segmentation accuracy over single point estimates as they produce regularized and smoother class labellings. Dollar et al. [31] proposed a method to enable the split of training samples $\mathcal{P}_i = (\mathcal{I}_i(\boldsymbol{p}), y_i(\boldsymbol{p}))$ with label patches $y \in \mathbb{Z}^{(M_e)^3}$ of size $(M_e)^3$. The label patches are mapped into an intermediate space in which the Euclidean distances between the samples can be used to perform dimensionality reduction. In this way, the label patches are mapped into one dimensional space and the standard entropy based training objective $\mathcal{H}_c$ can be used to split the training samples.

In a related work [33], this approach is applied on 3D cardiac MR and ultrasound images to generate probabilistic surface maps of the heart in medical images. Similarly, in this work, the leaf nodes cast predictions to find the probabilistic map of the object surface $E(\boldsymbol{p})$ for the given input channels.

$$p(E(\boldsymbol{p})\,|\,\mathcal{I}) = \frac{1}{|\Omega_{\boldsymbol{p}}| \cdot N_{\text{tree}}} \sum_{i \in \Omega_{\boldsymbol{p}}} \sum_{\ell \in \mathcal{L}_i} p(y(\boldsymbol{p}_i) = 1\,|\,\ell) \quad (3)$$

where $\Omega_{\boldsymbol{p}}$ is the set of input patches that are overlapping at the location $\boldsymbol{p}$. N_{tree} denotes the number of trees and $p(y(\boldsymbol{p}_i) = 1\,|\,\ell)$ is the probability of the event whether the binary edge patch stored in the leaf node ℓ is positive at $\boldsymbol{p}$ or not. The learned class posterior distributions in the leaf nodes are used as weighting terms in the regression function. The regression node training and inference information is provided in the next section.

C. Regression of Landmark Locations

The structured forest model can be enhanced by adding regression nodes in addition to the structured classification split nodes. In this way, the trained classifier can cast regression votes for anatomical landmark locations. More importantly, this combination enables to train class specific regression models. Each training sample $\{\mathcal{P}_i = (\mathcal{I}_i, y_i, \mathcal{D}_i)\}$ is now characterized by a set of offset vectors to each landmark point. The set of offsets $\mathcal{D}_i = (\boldsymbol{d}_i^1, \boldsymbol{d}_i^2, \cdots, \boldsymbol{d}_i^L)$ contain 3D displacement vectors from the patch centre to the target landmark location, where L is the number of landmarks.

The regression split nodes are trained by minimizing the determinant of full covariance matrix [19] defined by the landmark offset vectors. In this way, the inter-dependency between the landmark locations are partially taken into account, and it allows the model to learn an implicit shape model of the organ.

In the leaf nodes ℓ, the regression information from the training samples is stored using a parametric model $p(\boldsymbol{d}_\ell^n) = \mathcal{N}(\boldsymbol{d}^n; \overline{\boldsymbol{d}_\ell^n}, \Sigma_\ell^n)$: The mean $\overline{\boldsymbol{d}_\ell^n}$ and covariance matrix Σ_ℓ^n parameters are computed for the $3L$ dimensional multivariate normal distribution. This regression model is preferred over the non-parametric models, such as Parzen estimation or Mean-shift mode seeking, due to its computational simplicity. Similar to the work proposed in [21], offset vector models in the leaf nodes are conditioned based on segmentation label of the training samples. In other words, samples collected around the target organ surface have a separate regression model $p(\boldsymbol{d}^n|y = k, \ell)$ than the background samples. As in [10], [21], we assume that the background pixels are not as informative as the organ surface voxels in detecting anatomical landmarks.

As the pose variations can be quite large, long range regression votes are observed to be decreasing the landmark localization accuracy and confidence. Therefore, the Hough vote maps $F(\boldsymbol{p}^n)$ are formulated as follows:

$$p(F(\boldsymbol{p}^n)\,|\,\mathcal{I}) = \sum_{k=\{0,1\}} p(F(\boldsymbol{p}^n), y(\boldsymbol{p}) = k\,|\,\mathcal{I})$$

$$p(F(\boldsymbol{p}^n)\,|\,\mathcal{I}) \approx p(F(\boldsymbol{p}^n)\,|\,\mathcal{I}\ y(\boldsymbol{p}) = 1)\, p(y(\boldsymbol{p}) = 1) \quad (4)$$

Here only the points along the organ surface ($y(\boldsymbol{p}) = 1$) are allowed to cast votes for the landmark locations. However, landmarks can be positioned at any arbitrary location either inside or along the organ surface. The probability of a voxel classified as a point on the organ surface is obtained through the use of classification node splits (Section III-B). In this way, the landmark prediction function is constructed as:

$$\begin{aligned} &p(F(\boldsymbol{p}^n)\,|\,\mathcal{I}) \\ &\approx \frac{1}{N_{\text{tree}} \cdot N} \\ &\cdots \sum_{i=1}^{N} \sum_{\ell \in \mathcal{L}_i} w_\ell^n\, p(E(\boldsymbol{p}\,|\,\mathcal{I}))\, K\left(\frac{\boldsymbol{p}^n - \left(\boldsymbol{p}_{(i,c)} + \overline{\boldsymbol{d}_\ell^n}\right)}{h} \right) \end{aligned}$$
$$(5)$$

where N is the number of image voxels, K is a Gaussian kernel with bandwidth parameter h, $w_\ell^n = \frac{1}{trace(\Sigma_l^n)}$ is the confidence parameter for landmark n. The subscript (i, c) denotes the centre voxel of input patch. After computing the Hough vote map, the landmark location is determined by choosing the voxel with highest probability value as proposed in [19]:

$$\hat{\boldsymbol{p}}^n = \underset{\boldsymbol{p}^n}{\arg\max}\; p(F(\boldsymbol{p}^n)\,|\,\mathcal{I}) \quad (6)$$

D. Stratification of Global Characteristics

Modelling the offset vectors to landmarks in leaf nodes as Gaussian distribution biases the learning towards the average landmark distribution observed in training data. In datasets with smaller variations in terms of pose and object shape, this is unlikely to introduce too much error. However, in cases such as cardiac MR imaging, the orientation and size of the heart can exhibit large degrees of variation. For these cases, it is useful to have population groupings in sub-trees to increase the localization accuracy.

Along the same line, neighbourhood approximation forests [25] were proposed to cluster brain MR images of patients from different population groups. Our method takes this approach one step further by making use of the implicit clustering in a decision tree composed of classification and regression nodes. This clustering of the data can be viewed as a population stratification and allows our method to achieve improved landmark localization accuracy. The proposed method does not require additional training data and information compared to standard Hough forests.

To achieve this goal, the training images are annotated with size and pose parameters. These parameters are computed automatically using only the given input landmark point sets.
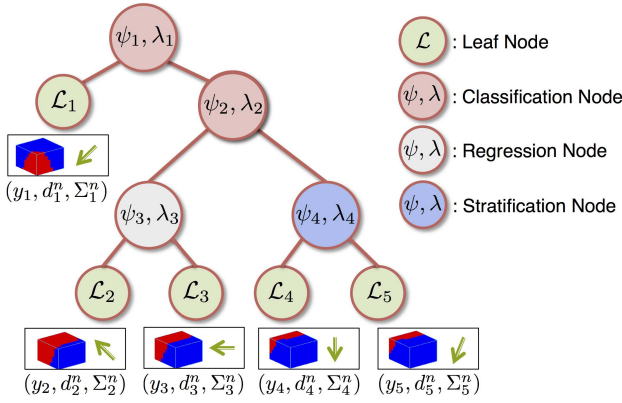
Fig. 2.   An illustration of the stratified decision tree structure. The standard Hough tree structure [10] is modified by adding the proposed stratification node splits.

A reference point set $\mathcal{F}_r = \{\boldsymbol{p}_r^1, \boldsymbol{p}_r^2, \cdots, \boldsymbol{p}_r^L\}$, obtained from a selected reference MR cardiac image, is used to quantify in-plane orientation θ_i and size β_i of a given training image. To obtain these parameters for each training image i, the ground-truth landmark point set $\mathcal{F}_i$ is aligned to $\mathcal{F}_r$ by computing an affine transformation $T_{(i,r)}$.

The meta information $\mathcal{M}_i = \{\theta_i, \beta_i\}$ is used in stratification nodes ψ_s to cluster training samples with similar pose and size. The impurity of training samples $\mathcal{S} = \{\mathcal{P}_i = (\mathcal{I}_i, y_i, \mathcal{D}_i, \mathcal{M}_i)\}$ in terms of the size parameter is defined based on mean squared differences $H_\beta(\mathcal{S}) = \sum_i (\beta_i - \tilde{\beta})^2$. Similarly we compute the impurity in terms of the orientation parameter $H_\theta(.)$. The two impurity measures H_β and H_θ may have quite different ranges depending on the problem and its dimensionality, and one of them could easily dominate the other one during optimization. Hence, we combine the two in a single stratification uncertainty term $H_s(.)$ by normalizing the two uncertainty terms similar to the joint training objective proposed in [34]:

$$H_s(\mathcal{S}) = \frac{1}{2}\left(\frac{H_\theta(\mathcal{S})}{H_\theta(\mathcal{S}_0)} + \frac{H_\beta(\mathcal{S})}{H_\beta(\mathcal{S}_0)}\right) \qquad (7)$$

Here $\mathcal{S}_0$ denotes the training sample set reaching the root node, and $\mathcal{S}$ represents the selected samples after the node split. With this formulation, we assume that the random variables corresponding to size and orientation are independent of each other, i.e. the joint entropy can be expressed as $p(\theta, \beta | \mathcal{I}) = p(\theta | \mathcal{I}) \cdot p(\beta | \mathcal{I})$. In the negative log domain, the joint entropy is expressed as summation of the two terms, which leads to the equation in (7).

As shown in Figure 2, the trees are trained in a joint manner using the following split node types: structured classification ψ_c, regression ψ_r, and stratification nodes ψ_s. As proposed in [10], the training objective at each split node is selected randomly among the listed three objectives. Based on our initial experiments, the node selection probabilities are fixed to $p(c) = p(r) = 0.4$, $p(s) = 0.2$.

The stratification splits can be trained with the box features presented in Section III-A. However, global scale features are found to be more expressive in separating images in terms

of shape and pose parameters as shown in the work [23] of iterative facial landmark localization. For this reason, we introduce global shape features to represent all samples collected from the same image, and they are used only for the stratification splits. These features are (i) inter-landmark distances and distance ratios [23], which are calculated through initial landmark location predictions, and (ii) histograms of gradients (HoG) [28] computed using only the initial edge map prediction of the target organ. Based on our experiments, these features achieve better stratification and regression results; however, their computation requires a two stage cascaded model as the landmark distance features are computed based on the initial estimates of the landmark locations. In other words, the initial tree model in the cascaded approach is used only to obtain a coarse representation of the organ surface and landmark locations, which are later used to perform stratification splits in training of subsequent decision trees.

E. Visualization of the Stratification Splits

To better understand the advantages of the proposed method, a proximity analysis is performed on a trained stratified decision forest. In the training procedure, this analysis is normally not required to be done; however, it is a useful technique to understand the role of the stratification splits. In more detail, we are able to visualize the mapping of the training images from the root node to leaf nodes. The computed proximities will demonstrate that images with similar organ size and pose parameters are automatically mapped to closer leaf nodes in the tree structure.

To perform this analysis, the trained forest is analysed by computing the proximity matrix of the training samples, as explained in [27], which is a $M \times M$ matrix and M is the total number of training images. If the two images are mapped into the same leaf node, then their proximity is increased by one. Similarly, an adjacency matrix can be derived from these proximities, and the connections between the images can be visualized by applying a non-linear dimensionality reduction technique. In our setting, the adjacency matrix is analysed using the Laplacian Eigenmaps method [35], and the images are mapped into a two dimensional manifold space as can be seen in Figure 3.

In this figure, it can be seen that the cardiac short axis images with similar in-plane orientation and heart size are mapped into closer leaf nodes and share the same sub-trees. In this way, regression and classification information stored in leaf nodes are population specific, and it allows the nodes to make more accurate landmark predictions.

IV. RESULTS

For the training and evaluation of the proposed method, we used two separate and disjoint cardiac MR datasets, which are referred to as $Dataset1$ and $Dataset2$. The first set contains 1080 cine short-axis MR images of resolution $1.50 \times 1.50 \times 2.00$ mm from the UK Digital Heart Project [36]. $Dataset1$ is randomly partitioned into three equal sized subsets for 3-fold cross-validation, so in total 720 images are used to train the models for all the methods.
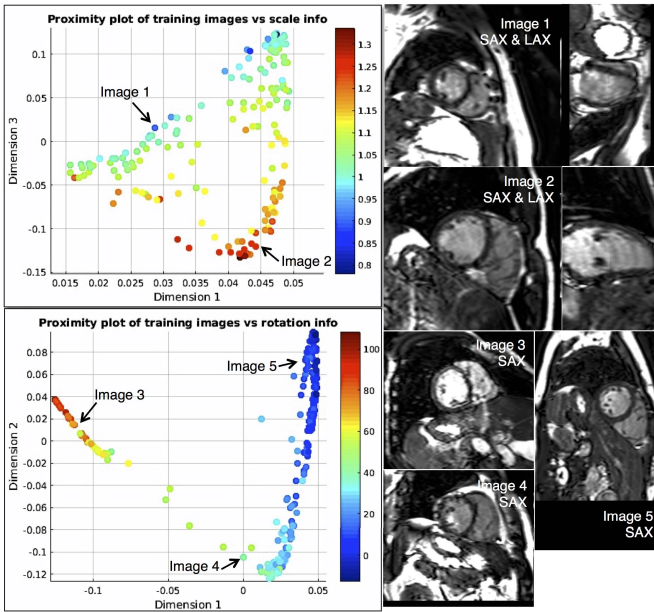
Fig. 3. Proximity plot (left) of the training images is shown, where the color code displays the heart orientation (bottom) and size (top) information of each image. The axes correspond to the first 3 dimensions in the lower dimensional space. The selected images are visualized on the right side to demonstrate the size and pose variations.
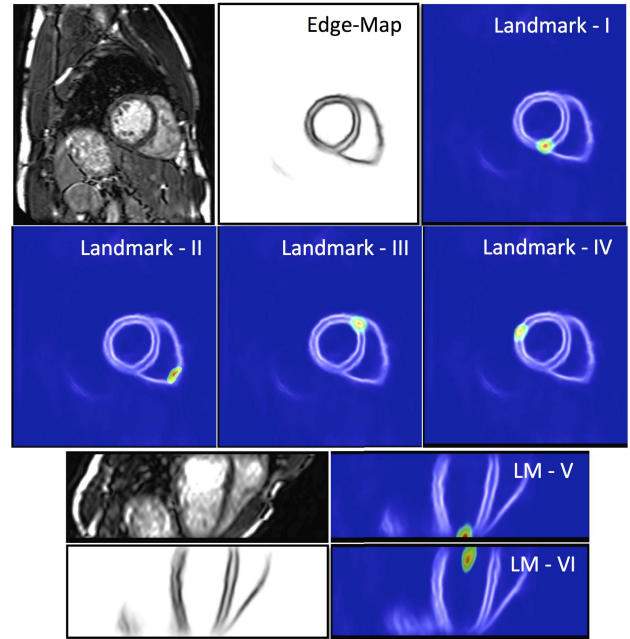

Fig. 4. From top left to bottom right: Input cardiac MR image, probabilistic surface map of the heart, and Hough vote maps for the location of six different anatomical landmarks. The Hough vote maps, shown in jet color-map, are obtained with the stratified forest, and they are overlaid on top of the probabilistic surface maps. Voxels with high probability are shown in red color.

The second dataset consists of 90 cine short-axis MR images with lower resolution $1.50 \times 1.50 \times 8.00$ mm. This dataset is part of the Cardiac Atlas Project (CAP) [37] and it is publicly available. The images from both datasets were acquired in different clinical sites with different MR scanners. As a preprocessing step, all images are linearly up-sampled to the same resolution as the images in the first dataset. Additionally, a data-augmentation strategy [38] is utilized to increase the number of training samples, which is performed by using label-preserving spatial transformations.

For evaluation of the proposed method, two different types of experiments are performed using these datasets. The first experiment demonstrates the accuracy of different landmark localization methods. The proposed stratified forest is compared against the standard Hough forest and image registration based localization techniques in two separate sub-experiments. For all the methods, we used the same training datasets and data-augmentation strategy. In the first experiment, the proposed method is compared to the baseline localization results obtained from the standard Hough forest. In the second experiment, image registration methods (block-matching [5] and 3D-SIFT alignment [6]) are employed prior to Hough forest to reduce the pose and size variations of the heart observed in the training and testing images. The spatially aligned images are later processed with Hough forest for training and inference purposes.

The second evaluation focuses on the application of the proposed method for image segmentation. A state-of-the-art multi-atlas segmentation method [26] is augmented with the proposed landmark localization method, and the obtained results are compared against the current semi-automatic approach in which the landmarks are identified manually.

A. Anatomical Landmark Localization

The decision trees for Hough and stratified forests are trained to locate $|L| = 6$ anatomical landmark locations. These landmark locations correspond to LV lateral wall mid-point, RV insert points (intersection between the RV outer boundary and the LV epicardium), RV lateral wall turning point (the point where the RV outer boundary changes directions significantly within the image), apex, and center of the mitral valve. The anatomical landmarks are shown in Figure 4. In Table II, the detection results for these landmark locations are shown, which are produced by averaging the results obtained from the three folds of the cross-validation. The localization errors are reported in terms of mean and median of Euclidean distance between the detected landmark position and the corresponding ground truth which is manually annotated by two experts. Additionally, the location of the center point of these landmarks $\hat{p}^c = \frac{1}{L} \sum_{n=1}^{L} \hat{p}^n$ is also computed as this point is less influenced by the inter-observer variability in the manual annotations. The experimental results show that the use of the proposed stratification split nodes significantly increases the landmark localization accuracy as it is better able to cope with variations of the size and pose of the heart as observed in the cardiac MR images. Table II also shows the distribution of the errors along each image axis. The errors are mostly concentrated in the through plane direction due to the lower resolution in that direction. In Table III, the localization errors for each landmark point are shown.

The predictions for the second landmark (RV lateral) have a higher error compared to the other landmarks. The significant performance difference is due to lack of consistent definition

TABLE II

LANDMARK LOCALIZATION ERRORS FOR THE TWO DATASETS, NAMELY *Dataset*1 (1.5 × 1.5 × 2.00 mm) AND *Dataset*2 (1.5 × 1.5 × 8.0 mm). THE LANDMARK LOCALIZATION ERRORS ARE REPORTED IN TERMS OF MEAN AND MEDIAN EUCLIDEAN DISTANCES FOR ALL 6 LANDMARK LOCATIONS. ADDITIONALLY, THE LOCALIZATION ERROR FOR THE CENTRE POINT OF ALL THE LANDMARKS IS PROVIDED IN THE CENTRAL COLUMN. THE LAST COLUMN SHOWS THE LOCALIZATION ERRORS IN EACH ORTHOGONAL DIRECTION IN THE IMAGE SPACE. THE PROPOSED STRATIFIED FOREST (D) IS COMPARED AGAINST THE 3D-SIFT ROBUST ALIGNMENT [6] & HOUGH FOREST (A), STANDARD HOUGH FOREST (B), AND BLOCK MATCHING [5] & HOUGH FOREST (C) APPROACHES. PLEASE REFER TO SECTION IV FOR A DETAILED DESCRIPTION OF THE APPROACHES. THE INTER-OBSERVER ERRORS ARE REPORTED IN (E)

		Individual Landmarks, $|L| = 6$		Centre Point of Landmarks		Distribution of the Errors		
		Mean (mm)	Med (mm)	Mean (mm)	Med (mm)	X (mm)	Y (mm)	Z(mm)
Dataset1	(A)	14.51±10.74	12.236	6.92±5.33	5.187	6.53±6.83	6.47±6.62	7.70±7.64
	(B)	10.74±6.32	9.224	6.05±4.12	5.572	5.36±4.38	5.47±4.45	5.86±4.89
	(C)	10.44±8.24	8.399	5.33±4.30	4.283	4.21±4.92	4.46±5.14	5.22±6.41
	(D)	8.03±5.11	7.067	4.10±2.81	3.422	3.37±3.78	3.31±3.73	3.65±4.13
	(E)	6.67±4.17	6.293	2.90±3.42	2.842	2.44±3.61	2.40±3.94	2.99±5.50
Dataset2	(A)	16.36±10.35	14.798	8.48±6.82	6.928	7.72±8.09	6.83±7.09	8.07±6.65
	(B)	12.42±6.12	11.245	7.36±3.61	6.314	5.75±4.23	5.62±4.23	7.65±5.64
	(C)	12.54±10.25	10.587	7.10±6.36	5.385	5.86±8.04	5.19±5.20	6.20±6.99
	(D)	9.12±4.64	8.412	4.62±2.63	4.312	3.65±3.31	3.51±2.13	5.43±5.71
	(E)	6.82±3.95	6.197	2.96±2.04	2.719	2.54±2.74	2.57±3.42	3.03±3.16

TABLE III

LANDMARK LOCALIZATION ERRORS FOR THE TWO DATASETS, NAMELY *Dataset*1 (1.5 × 1.5 × 2.00 mm) AND *Dataset*2 (1.5 × 1.5 × 8.0 mm). THE LOCALIZATION ERRORS FOR EACH LANDMARK POINT IS PROVIDED IN TERMS OF MEAN EUCLIDEAN DISTANCE. THE PROPOSED STRATIFIED FOREST (D) IS COMPARED AGAINST THE 3D-SIFT ROBUST ALIGNMENT [6] & HOUGH FOREST (A), STANDARD HOUGH FOREST (B), AND BLOCK MATCHING [5] & HOUGH FOREST (C) APPROACHES. PLEASE REFER TO SECTION IV FOR A DETAILED DESCRIPTION OF THE APPROACHES. THE INTER-OBSERVER ERRORS ARE REPORTED IN (E)

		LM1 (mm)	LM2 (mm)	LM3 (mm)	LM4 (mm)	LM5 (mm)	LM6 (mm)	P-val
Dataset1	(A)	11.14±7.62	21.90±13.84	13.00±9.71	13.64±7.68	16.49±11.76	10.94±7.88	-
	(B)	9.71±5.65	17.03±8.88	10.65±5.20	10.18±5.77	8.62±4.55	8.27±4.51	-
	(C)	8.80±6.73	18.00±16.73	9.51±6.75	10.62±8.26	7.99±5.71	7.70±5.25	-
	(D)	7.02±4.93	12.23±6.94	7.92±5.52	7.36±4.22	6.74±4.12	6.32±3.95	-
	(E)	5.88±4.65	10.60±8.09	5.60±5.79	6.67±5.30	5.97±3.28	5.30±2.98	-
	(C-D)	1.78	5.77	1.59	3.26	1.25	1.38	≪1e-3
Dataset2	(A)	13.28±8.60	21.96±9.70	12.95±7.11	15.92±8.75	19.06±8.53	15.02±13.47	-
	(B)	12.12±6.87	19.03±12.31	10.41±6.64	12.31±5.94	11.31±4.21	9.13±4.53	-
	(C)	11.78±11.04	18.76±19.76	10.16±6.83	13.14±12.18	12.27±6.30	9.15±5.47	-
	(D)	8.42±4.42	10.12±5.02	8.04±4.08	11.32±4.16	10.41±4.15	7.21±3.38	-
	(E)	6.99±4.40	8.34±4.52	6.78±4.37	6.66±3.91	6.23±2.89	5.96±3.61	-
	(C-D)	3.36	8.64	2.12	1.82	1.86	1.93	≪1e-3

of RV lateral landmark, and the large shape variability of RV wall. These factors also increase the variance in manual annotations for this particular landmark. Compared to the Hough forest based localization, the proposed approach improves the detection accuracy significantly in both of the datasets. To observe the number of failure cases, the distribution of the errors is shown in the histogram in Figure 5. One can see that there are only a few outliers in the histogram. Moreover, the proposed method consistently performs better than the Hough forest based method as it can be seen on the cumulative distribution of the errors.

It is observed that a slight performance improvement in mean accuracy can be achieved when the Hough forest is preceded by a robust image alignment method. However, this improvement comes at the cost of increased variance of errors when the registration algorithm fails to align images. This situation can be explained by two reasons: (i) Variation in the training data is reduced due to spatial alignment of images to a reference space. (ii) Image alignment results are susceptible to large anatomical variations observed in organs other than heart, which leads to incorrect alignment results as the heart label information is not used. In particular, the poor registration results obtained with 3D SIFT features are attributed to the large slice thickness, which reduces the number of matched features between the images.

In addition to the landmark localization results, the performance of the pose and size estimations is measured. The root mean square (RMS) is adopted to evaluate the pose
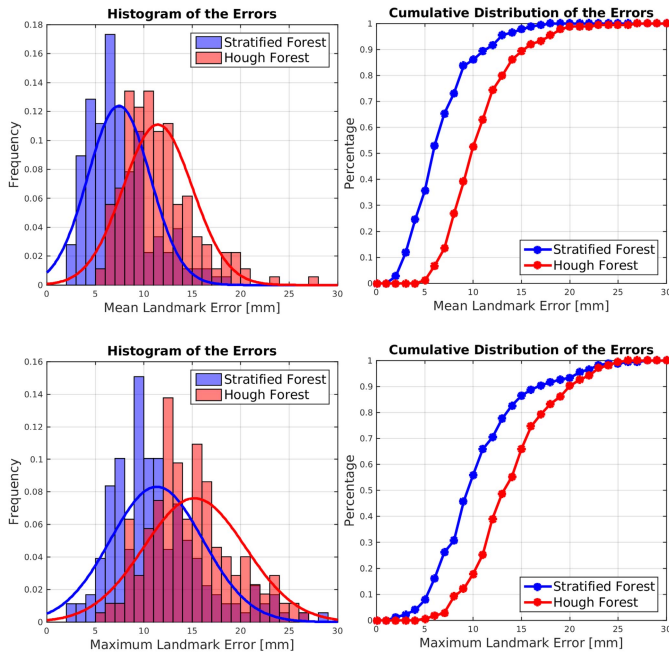
Fig. 5. Histogram of the landmark localization errors for the *Dataset*1. The distribution of mean (top) and maximum (bottom) localization errors are shown.

TABLE IV

HEART POSE AND SIZE ESTIMATION ACCURACY RESULTS FOR THE *Dataset*1 (*DS*1) AND *Dataset*2 (*DS*2). DISTRIBUTION OF THE GROUND-TRUTH POSE VALUES ARE REPORTED IN TERMS OF VALUE RANGE AND STANDARD DEVIATION. THE ROTATION VALUES ARE GIVEN IN RADIANS AND SCALE VALUES ARE UNITLESS

Dataset and pose parameter		Ground-truth value range	Standard deviation	Estimation error (RMS)
$DS1$	Rot	[-0.24 1.81]	0.501	0.11±0.08
	Size	[0.86 1.33]	0.097	0.05±0.03
$DS2$	Rot	[-0.36 1.67]	0.447	0.12±0.11
	Size	[0.91 1.32]	0.083	0.05±0.04

estimation results, and the ground truth information is obtained by globally aligning the manually annotated landmark points with an affine transformation. The pose estimation results for both datasets are shown in Table IV. The results show that the proposed stratified forest estimates the target pose and size parameters very accurately, which results in improved landmark localization performance.

B. Cardiac MR Image Segmentation

Landmark localization can be useful for subsequent image analysis such as segmentation and registration. In our experiments, a state-of-the-art multi-atlas patch based segmentation technique [39] is selected for this task. In comparison to classifier based methods, multi-atlas approaches have been more successful in the semantic segmentation of cardiac MR images as reported in the recent RV segmentation challenge [26]; however, they often require manual initializations. For this reason, the proposed landmark localization is used in a

multi-atlas segmentation framework to create a fully-automatic segmentation pipeline.

The localized landmarks are used to initialize the registration algorithm between the target and atlas images. A similarity transformation with 9 degrees of freedom is computed using the detected landmark point sets. The global alignment is followed by a B-spline based free-form deformable image registration [8]. The propagated labels are fused together using majority voting, and a graph-cut segmentation is applied as a post-processing to fill the gaps and smooth the segmentation labels.

In the experiments, four different segmentation frameworks are tested. All the frameworks made use of the same multi-atlas segmentation method, but the registration initializations were done either with different landmark point sets or by using an affine image registration. In that respect, we used manual annotations (I), Hough forest based landmarks (II), the proposed stratified decision forest based landmarks (III), and a robust block matching method (IV) [5]. In this experiment, we investigate the influence of landmark localization accuracy on segmentation results. Also, we compare the performance difference between the landmark and affine registration based initialization [40] on the segmentation method. Here, a separate dataset of cardiac images is used, which is publicly available and part of the Cardiac Atlas Project [37]. The dataset was used in the MICCAI'13 SATA challenge to benchmark different cardiac segmentation methods. It consists of 50 manually annotated image sequences acquired from healthy subjects, patients with LV hypertrophy and wall abnormalities due to prior myocardial infarction. Segmentations are performed only on the end-diastolic frames extracted from the sequences, and RV segmentation labels are collected by manually annotating the images. The segmentation results, shown in Figure 6, are achieved using 20 fixed atlases selected from the dataset.

Based on the results, we observe that the robust block matching method (RBM), in some cases, completely fails to find correspondences between the atlas and target images, which causes the observed outliers and large variations in the segmentation accuracy. It can be mainly explained with orientation and shape differences between the target and atlas images, and RBM is more sensitive to these conditions. On the other hand, the results show that most of the landmark localization errors can be compensated during the segmentation process. For this reason, the performance difference between the two automatic localization methods is less prominent in the segmentation results. However, we observe that in some cases the Hough forest landmark predictions fail, and this results in incorrect segmentations as can be seen in Figure 6 in form of outliers. More importantly, it is observed that the proposed landmark localization and segmentation initialization method yields similar accuracy as the semi-automatic segmentation method that is based on manual landmark identification. Therefore, we conclude that the proposed method is accurate enough in landmark localization to guide a multi-atlas segmentation method.

In a different experimental setting, the stratified forest and block matching methods were applied together sequentially to initialize the multi-atlas segmentation. This approach did
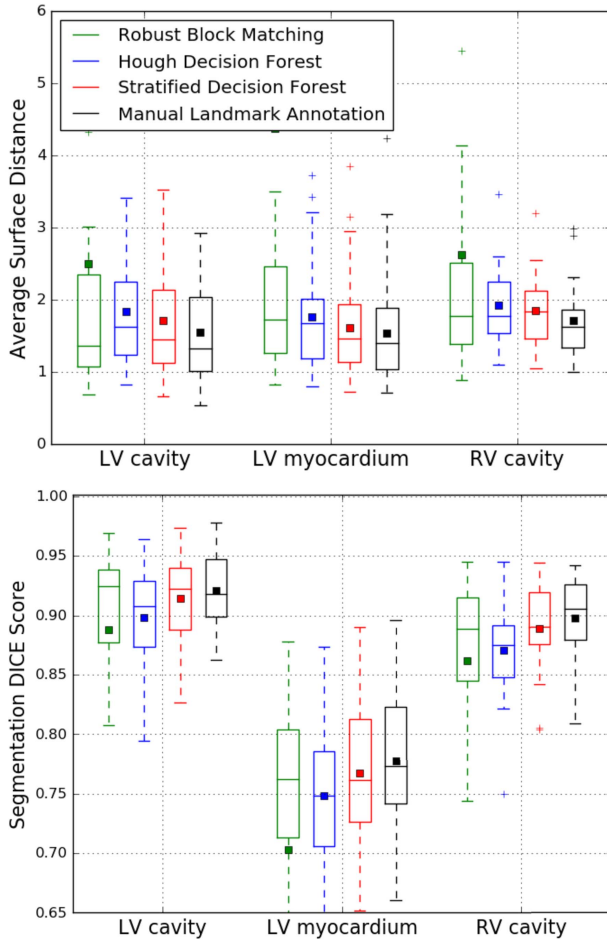
Fig. 6. Cardiac MR image multi-atlas segmentation results: dice score and mean surface distances are obtained for four different registration initialization techniques. Mean values are shown with coloured square boxes.

TABLE V
STRATIFICATION FOREST PARAMETER SPECIFICATIONS

Parameter value	Description
$(M_a)^3 = (28)^3$	3D Input patch size
$(M_e)^3 = (12)^3$	3D Output PEM patch size
$N_{\text{tree}} = 10$	Number of decision trees
$L = 6$	Number of landmarks
$D_{\max} = 36$	Maximum allowed tree depth
$\alpha = 7$	Number of feature channels 3 channels for stratification splits
$N_{\text{pos}} = 1.5 \times 10^6$	Number of positive training samples
$N_{\text{neg}} = 1.5 \times 10^6$	Number of negative training samples

TABLE VI
LOCALIZATION ACCURACY OF RV INSERTION POINTS ON 2D SHORT AXIS SLICES. THE PROPOSED STRATIFIED FOREST IS COMPARED AGAINST THE BOOSTING TREE CLASSIFIER BASED LANDMARK LOCALIZATION METHOD [41] IN TERMS OF COMPUTATION TIME AND LOCALIZATION ACCURACY. THE METHODS ARE BENCHMARKED ON THE SAME DATASET ($Dataset2$) THAT WAS USED IN THE STACOM'12 LV LANDMARK DETECTION CHALLENGE. THE NAME OF THE DATASETS ARE ABBREVIATED AS $DS1$ AND $DS2$ IN THE TABLE

Evaluation Dataset and Method		Mean±Std (mm)	Median (mm)	Time (sec)
$DS1$	Stratified Forest	4.63±3.81	3.57	0.11
$DS2$	MSL PBT [27]	7.90±11.50	4.70	0.14
	Stratified Forest	5.55±3.66	4.42	0.11

not improve the segmentation accuracy significantly. However, as the computational cost increase significantly with block matching, in practice one can safely opt for a stratified forest only based initialization.

V. IMPLEMENTATION DETAILS

In Table V, the parameter values used in the experiments are specified. Better localization results could be obtained on the same datasets by tuning the parameter values, which was not performed in our experiments. As the leaf nodes store structured patch label information, the number of required randomized trees is not as large as the standard decision forest approaches. Landmark localization and block matching experiments were carried out on a Intel-i7 3.40 GHz quad-core machine, and the approximate computation time to initialize each atlas was 12s for stratified forest and 2.1min for block matching. The non-rigid registration experiments were conducted on a machine with a graphical processing unit (GPU), and the average computation time was 49.2s per atlas. Therefore, with the proposed fully-automatic segmentation pipeline each atlas can be accurately mapped into target image space within a minute.

VI. DISCUSSION

The proposed stratified forest method is compared with the MSL classifier based landmark localization approaches [11], [12]. In [41], RV insert points localization results on 2D short axis images are reported, which were evaluated on the same CAP testing dataset as the one used in our experiments ($Dataset2$). The results presented in Section IV-A are recomputed in 2D space by projecting the ground-truth and detected landmark points to the corresponding short axis plane. The corresponding landmark localization results and computation time per 2D slice are shown in Table VI.

Similarly, the regression only forest is tested on the same image datasets, and the localization results are observed to be less accurate than the Hough forest based localization results. The difference can be explained with the fact that in Hough forest landmark votes are collected only from the heart surface, and this produces more consistent landmark position hypotheses. However, regression only approaches have been used in medical image analysis, particularly in [19], [20] to detect bounding boxes around the abdominal organs in CT images.

Other multi-atlas label fusion techniques [42], [43] could be chosen to segment cardiac MR images. These methods showed

slight performance improvement over conventional majority voting and patch based fusion techniques when they were used to segment brain MR [42] and abdominal CT [43] images. In particular, the key-point transfer segmentation method eliminates the need for non-rigid image registration between target and atlas images, which basically reduces the dependency of segmentation algorithm on prior image alignment accuracy.

Moreover, in the experiments, a graph based regularization is tested on Hough vote images (in total 6 channels) as a post-processing step to constrain the landmark detections based on a learned inter-landmark distance model. However, we have not observed an improved performance by taking this approach. This can be explained with the joint training of all landmarks, where the regression training objective minimizes the determinant of covariance matrix in a joint manner for all the landmarks. Therefore, our approach does not require this type of post-processing as in [14].

During the inference, the proposed stratified forest does not predict a single deterministic orientation and scale value as in the case of the cascaded pose estimation methods. Testing samples can be routed in leaf nodes with different pose parameter models, which enables a probabilistic modelling of the latent variable. This can be considered as another improvement over the marginal space learning based approaches.

In the experiments, we performed additional tests by allowing the background long-range pixels to cast Hough votes as well. It is observed that due to pose and shape variations, the background pixels introduce more dispersed landmark prediction maps which reduces the algorithm accuracy. For this reason, the predictions are only performed by the heart surface voxels in the images. Additionally, one could learn regression models conditioned on the label of different heart tissues. However, this would require multi-class segmentation.

VII. CONCLUSION

In this paper, a novel learning objective is proposed to learn more representative and patient specific decision forest classifier and regressor models. This new feature increases the landmark localization accuracy, as the models trained with stratification splits are better able to cope with pose and size variations of the organs observed in the images. Moreover, the proposed method provides better guidance for the subsequent image analysis techniques. As shown in the experiments, state-of-the-art multi-atlas segmentation achieves better accuracy and displays robust performance when the proposed method is used an initialization technique. Moreover, the proposed patient stratification approach is generic and modular; as such it can be used in any decision tree structure to achieve better classification and regression results. This includes applications to different modalities and other target organs. In that regard, the future work will investigate the use of stratified decision forests on 3D ultrasound images to identify viewing planes and organ locations.

REFERENCES

[1] W. Bai *et al.*, "A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: Application to cardiac MR images," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1302–1315, Jul. 2013.

[2] X. Lu *et al.*, "Automatic view planning for cardiac MRI acquisition," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Springer, 2011, pp. 479–486.

[3] M. D. Cerqueira *et al.*, "Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart a statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association," *Circulat.*, vol. 105, no. 4, pp. 539–542, 2002.

[4] K. Babalola, A. Gait, and T. F. Cootes, "A parts-and-geometry initialiser for 3D non-rigid registration using features derived from spin images," *Neurocomputing*, vol. 120, pp. 113–120, Nov. 2013.

[5] S. Ourselin, A. Roche, S. Prima, and N. Ayache, "Block matching: A general framework to improve robustness of rigid registration of medical images," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Springer, 2000, pp. 557–566.

[6] M. Toews and W. M. Wells, "Efficient and robust model-to-image alignment using 3D scale-invariant features," *Med. Image Anal.*, vol. 17, no. 3, pp. 271–282, 2013.

[7] T. F. Cootes, C. J. Twining, V. S. Petrovic, K. O. Babalola, and C. J. Taylor, "Computing accurate correspondences across groups of images," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 11, pp. 1994–2005, Nov. 2010.

[8] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[9] V. Potesil, T. Kadir, G. Platsch, and M. Brady, "Personalized graphical models for anatomical landmark localization in whole-body medical images," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 29–49, 2015.

[10] J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky, "Hough forests for object detection, tracking, and action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2188–2202, Nov. 2011.

[11] Y. Zheng, X. Lu, B. Georgescu, A. Littmann, E. Mueller, and D. Comaniciu, "Robust object detection using marginal space learning and ranking-based multi-detector aggregation: Application to left ventricle detection in 2D MRI images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1343–1350.

[12] X. Lu *et al.*, "Cardiac anchoring in MRI through context modeling," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, Springer, 2010, pp. 383–390.

[13] J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 1297–1304.

[14] R. Donner, B. H. Menze, H. Bischof, and G. Langs, "Global localization of 3D anatomical structures by pre-filtered Hough Forests and discrete optimization," *Med. image Anal.*, vol. 17, no. 8, pp. 1304–1314, 2013.

[15] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, and D. Comaniciu, "3D deep learning for efficient and robust landmark detection in volumetric data," in *Medical Image Computing Computer-Assisted Intervent (MICCAI)*. Springer, 2015, pp. 565–572.

[16] T. Nelson, D. Pretorius, A. Hull, M. Riccabona, M. Sklansky, and G. James, "Sources and impact of artifacts on clinical three-dimensional ultrasound imaging," *Ultrasound Obstetrics Gynecol.*, vol. 16, no. 4, pp. 374–383, 2000.

[17] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, "Efficient regression of general-activity human poses from depth images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 415–422.

[18] M. Sun, P. Kohli, and J. Shotton, "Conditional regression forests for human pose estimation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Jun. 2012, pp. 3394–3401.

[19] A. Criminisi *et al.*, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Med. Image Anal.*, vol. 17, no. 8, pp. 1293–1303, 2013.

[20] R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch, "Multi-organ localization combining global-to-local regression and confidence maps," in *Medical Image Computing Computer-Assisted Intervent (MICCAI)*. Springer, 2014, pp. 337–344.

[21] S. Schulter, C. Leistner, P. Wohlhart, P. M. Roth, and H. Bischof, "Accurate object detection with joint classification-regression random forests," in *Porc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 923–930.

[22] M. Dantone, J. Gall, G. Fanelli, and L. Van Gool, "Real-time facial feature detection using conditional regression forests," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog. (CVPR)*, Jun. 2012, pp. 85–2578.

[23] X. Zhao, T.-K. Kim, and W. Luo, "Unified face analysis by iterative multi-output random forests," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Jun 2014, pp. 1765–1772.

[24] H. Lombaert, D. Zikic, A. Criminisi, and N. Ayache, "Laplacian Forests: Semantic image segmentation by guided bagging," in *Medical Image Computing Computer-Assisted Intervent (MICCAI)*. Springer, 2014, pp. 496–504.

[25] E. Konukoglu, B. Glocker, D. Zikic, and A. Criminisi, "Neighbourhood approximation using randomized forests," *Med. Image Anal.*, vol. 17, no. 7, pp. 790–804, Oct. 2013.

[26] C. Petitjean *et al.*, "Right ventricle segmentation from cardiac MRI: A collation study," *Med. image Anal.*, vol. 19, no. 1, pp. 187–202, 2015.

[27] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[28] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. no. 1, Jun. 2005, pp. 886–893.

[29] P. Dollar, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. BMVC*, 2009, pp. 91.1–91.11.

[30] A. Criminisi, J. Shotton, and S. Bucciarelli, "Decision forests with long-range spatial context for organ localization in CT volumes," in *Proc. MICCAI Workshop Probabilistic Models Med. Image Anal. (PMMIA)*, Sep. 2009.

[31] P. Dollar and C. L. Zitnick, "Structured forests for fast edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1841–1848.

[32] P. Kontschieder, S. Rota Bulo, H. Bischof, and M. Pelillo, "Structured class-labels in random forests for semantic image labelling," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2190–2197.

[33] O. Oktay *et al.*, "Structured decision forests for multi-modal ultrasound image registration," in *Proc. Medical Image Computing Computer-Assisted Intervent (MICCAI)*. Springer, 2015, pp. 363–371.

[34] B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi, "Joint classification-regression forests for spatially structured multi-object segmentation," in *IEEE Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2012, pp. 870–881.

[35] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[36] W. Bai *et al.* "A bi-ventricular cardiac atlas built from 1000 + high resolution MR images of healthy subjects and an analysis of shape and motion," *Med. Image Anal.*, vol. 26, no. 1, pp. 133–145, 2015.

[37] C. G. Fonseca *et al.*, "The cardiac atlas project an imaging database for computational modeling and statistical atlases of the heart," *Bioinformatics*, vol. 27, no. 16, pp. 2288–2295, 2011.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Proc. Syst.*, 2012, pp. 1097–1105.

[39] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.

[40] M. A. Zuluaga, M. J. Cardoso, M. Modat, and S. Ourselin, "Multi-atlas propagation whole heart segmentation from MRI and CTA using a local normalised correlation coefficient criterion," in *Functional Imaging and Modeling of the Heart*. Springer, 2013, pp. 174–181.

[41] X. Lu and M.-P. Jolly, "Discriminative context modeling using auxiliary markers for LV landmark detection from a single MR image," in *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*. Springer, 2013, pp. 105–114.

[42] A. J. Asman, Y. Huo, A. J. Plassard, and B. A. Landman, "Multi-atlas learner fusion: An efficient segmentation approach for large-scale data," *Med. Image Anal.*, vol. 26, no. 1, pp. 82–91, 2015.

[43] C. Wachinger, M. Toews, G. Langs, W. Wells, and P. Golland, "Keypoint transfer segmentation," in *Information Processing in Medical Imaging*. Springer, 2015, pp. 233–245.