

ShapeForest: Building Constrained Statistical Shape Models with Decision Trees

Saša Grbić, Joshua K.Y. Swee, and Razvan Ionasec

Siemens Corporation, Corporate Technology, Princeton NJ, USA

Abstract. Constrained local models (CLM) are frequently used to locate points on deformable objects. They usually consist of feature response images, defining the local update of object points and a shape prior used to regularize the final shape. Due to the complex shape variation within an object class this is a challenging problem. However in many segmentation tasks a simpler object representation is available in form of sparse landmarks which can be reliably detected from images. In this work we propose ShapeForest, a novel shape representation which is able to model complex shape variation, preserves local shape information and incorporates prior knowledge during shape space inference. Based on a sparse landmark representation associated with each shape the ShapeForest, trained using decision trees and geometric features, selects a subset of relevant shapes to construct an instance specific parametric shape model. Hereby the ShapeForest learns the association between the geometric features and shape variability. During testing, based on the estimated sparse landmark representation a constrained shape space is constructed and used for shape initialization and regularization during the iterative shape refinement within the CLM framework. We demonstrate the effectiveness of our approach on a set of medical segmentation problems where our database contains complex morphological and pathological variations of several anatomical structures.

1 Introduction

The ability to delineate deformable objects from images is critical for many computer vision tasks. In most cases low-level information based on the local image appearance is combined with high-level information based on a shape model to estimate the final object in the image. The low-level information can be noisy due to missing or incomplete image gradients. Thus the shape model is used to regularize the results. The success of these approaches is highly dependent on the way shape priors are modeled, how the low-level information is computed from the images and what optimization approach is utilized to fit the shape model to the low-level information. One of the first approaches was the Active Contour Model proposed by Kass et al. [13]. Hereby, the shape prior is represented as a general regularity term during the optimization, assuming that the shape should deform like a thin plate. Further, more object-specific shape priors

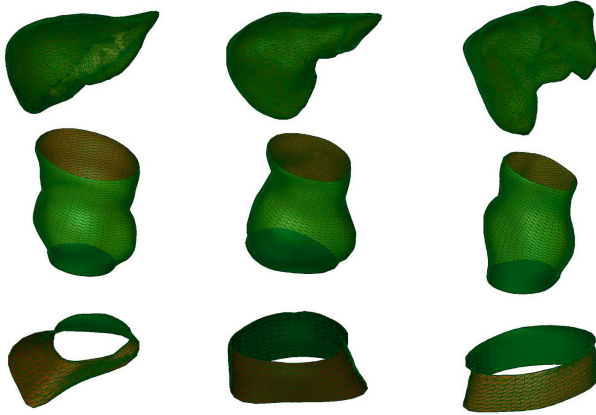


Fig. 1. Examples of complex variation observed among shapes within an object class for the liver (top row), the AV (middle row) and the LVOT (bottom row)

became prevalent, where the shape priors are learned from a set of training samples, such as Active Shape Models (ASM) [2]. Many extensions of this method were proposed within the last decade.

In current literature, shape models confront major challenges. First, shape variations are usually complex and therefore difficult to model using standard linear parametric models. Figure 1 illustrates the complex shape variation within our medical data set, showing three examples of the liver (top row), the aortic valve – AV (middle row) and the left ventricle outflow tract – LVOT (bottom row). Second, shape variation is often local. Shape models should be able to preserve local shape detail in the training data even if it is not statistically significant. Third, even though the final shape consists of hundreds or thousands of vertices a lower level representation using a small number of sparse landmarks associated with each shape is often available. These points define salient locations in the image and are located on or around the object of interest. These landmarks can be robustly detected using object detection algorithms. Shape models should exploit this information during shape alignment and refinement. r

In this paper we introduce a novel constrained statistical shape model, the ShapeForest, which addresses the main challenges presented above in a unified framework. Based on a set of sparse landmarks, which we assume can be detected reliably in the images, we infer an instance specific statistical shape model used for shape initialization and during the iterative shape refinement. We use randomized decision trees to learn the shape-manifold based on geometric features defined by the sparse landmark representation. During testing a subset of shapes is selected and used to construct a constrained parametric statistical shape model. We show that this approach can approximate accurately the shape variation compared to the classical statistical shape model and other multi-modal shape representations. We demonstrate the performance of the ShapeForest in a range of medical image segmentation problems – the aortic

valve, the left ventricle outflow tract, and the liver. In addition we show the effects of different choices for the training parameters.

2 Related Work

There is a wide range of literature on matching statistical shape models to images starting with Active Shape Models [2]. Hereby an iterative approach is pursued with an alternating strategy. First an update for each point is found in a small neighborhood of the current shape estimate with a suitably trained detector. Then the shape model parameters are adjusted to match the updated shape points. Active Appearance Models [3] align both models of shape and texture using an efficient parameter update schema. Pictorial structures define a collection of parts arranged in a deformable geometric configuration representing an object model [9]. Each part has an Active Appearance Model representation and the deformable geometric configuration is represented as spring-like connections between pairs of parts. An efficient method was introduced to fit the pictorial structure to an image. Constrained Local Models (CLM) [16,6,15] compute response images, measuring the fit of a specific shape point throughout the whole image. Based on the response images a shape model is matched to the data, estimating the best overall combination of points based on geometric criteria.

In the context of shape modeling, extensions to the classical SSM have been proposed to model multi-modal distributions within the shape prior. A simple extension is to use a mixture of Gaussian to represent the shape model [4]. Manifold learning can also be used to represent the non-linear shape prior [8,18,14]. A data-base guided approach was proposed in [11] whereby a nearest-neighborhood search is used to find the closest shape based on a data-base of shapes. However in most cases the number of clusters is defined a priori.

Another difficulty with SSM is the inability to preserve local details of the input shape when such details are present in the training data but not statistically significant. As PCA performs eigenanalysis and extracts eigenvectors with the largest eigenvalues the discarded eigenvectors are statistically insignificant, but they may contain important local details. Some relevant work can alleviate this problem. Sparse PCA obtains sparser modes and produces near-orthogonal components [17,19]. Thus each mode only affects locally clustered landmarks and captures more detail information. However in most sparse methods the number of training samples used for the sparse shape representation, especially with shapes which contain large number of vertices, is limited as it increases significantly the computational complexity. Some other methods divide the shape model into several independently modeled parts [7], such as the hierarchical approach. Since the smaller parts exhibit less variation, they can be captured with fewer training samples than the variations for the full shape.

Most existing shape segmentation methods pursue a hierarchical estimation approach. Based on lower level information such as a set of sparse landmarks a shape model is initialized and refined to fit the image data. These landmarks represent salient, corner-like structures in the images and are located either on

the shape or in close proximity. In our medical imaging domain the problem is further simplified as only one occurrence of each landmark can appear in an image. Given a large database of landmark annotations, state-of-the-art detection algorithms can be trained to detect those points reliably. A sliding window approach using discriminative classifiers with descriptive image features is often used in such settings [20]. Regression based approaches could also be utilized to detect the landmarks by exploiting the spatial correlation of image features and landmark positions [10,5]. In our work we assume this workflow as given.

Our method automatically selects a subset of relevant shapes from the training data for a new instance based on geometric features extracted from the detected landmarks. Thus there is no need to define a priori the number of clusters within our database of shapes. In addition the ShapeForest will learn a distance function between the landmarks and the remaining shape variation. Our approach is more robust compared to an Euclidean distance norm in the feature space as it discards non-informative and noisy features. The ShapeForest will select the most relevant geometric features to minimize the variance within the remaining subset of shapes. Thus the shapes that are selected will share geometric properties such as local geometric detail.

3 Shape Modeling

Given a population of n instances of a shape $S = \{s_1, s_2, \dots, s_n\}$ that have been aligned to a common coordinate system using Generalized Procrustes Analysis (GPA) and share point correspondences, each aligned shape instance s_n can be represented as a set of k points $s_n = \{s_n^{x1}, s_n^{y1}, s_n^{z1}, \dots, s_n^{xk}, s_n^{yk}, s_n^{zk}\}$. S can be viewed as a point cloud in $3k$ -dimensional space. A generative shape model, represented as a linear Gaussian parametric model, can be constructed by computing the Principal Component Analysis (PCA) on S . Later each shape s_u can be approximated using the following linear model:

$$s_u = T(\bar{s} + Pb) \quad (1)$$

Here, \bar{s} is the mean of all shapes in S after GPA, Pb is the linear uni-modal Gaussian parametric model with the eigenvectors P that represent the main modes of variation in S and the coefficients b , and T is a global transformation that maps \bar{s} to the coordinate system of s_u . Hereafter we will refer to this model as the *global SSM* or *gSSM* [12].

The ShapeForest instead identifies a subset of shape instances \hat{S} within S that share similar geometric characteristics based on geometric features extracted from an image I . The subset of shapes \hat{S} is later used to generate a constrained version of the global SSM:

$$s_u = T(\hat{s} + \hat{P}\hat{b}) \quad (2)$$

where \hat{s} is the mean shape computed from a subset of shape instances $\hat{S} \subseteq S$ after GPA, and $\hat{P}\hat{b}$ represents the corresponding modes of variation. We will later refer to this shape model as the *constrained SSM* or *cSSM*.

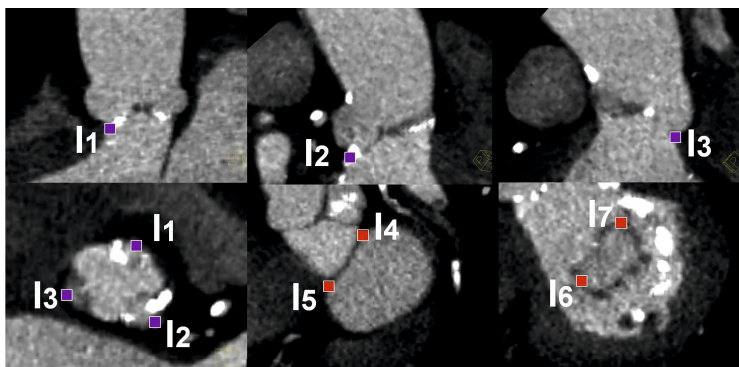


Fig. 2. Sparse Landmark representation $L = \{l_1, l_2, l_3, l_4, l_5, l_6, l_7\}$ for the left ventricle outflow tract (LVOT) in a 3D CT image. Each landmark represents a salient location in the image.

3.1 Landmarks

To select \hat{S} we use geometric features derived from m sparse landmarks L which are detected from the image I . Each landmark corresponds to a salient anatomical point in the image I and can be easily identified and reliably detected using standard object detection algorithms. Landmarks do not need to lie on the shape itself but can also be in close proximity to it. It is important to note that $m \ll k$. Figure 2 illustrates seven sparse landmarks associated with a shape (LVOT). Here, 5 landmarks lie directly on the shape itself, while the remaining 2 represent nearby stable points.

3.2 Geometric Features

From the identified m landmarks L , we compute two simple geometric features: *distance* features and *random plane* features. As the variability of the landmarks is strongly correlated with the shape points, these features are powerful to characterize the morphological and geometric properties of individual shapes.

Given a set of m landmarks associated with a shape, distance features (f_{dist}) are generated for each unique pair of landmarks (p, q) as the Euclidean distance between landmarks p and q :

$$f_{dist}(p, q) = \sqrt{\sum_{i \in \{x, y, z\}} (p^i - q^i)^2} \tag{3}$$

By comparison, random plane features (f_{rp}) are generated for each individual landmark p as the shortest distance between landmark p and a randomly generated plane:

$$f_{rp}(p) = \frac{ap^x + bp^y + cp^z + d}{\sqrt{a^2 + b^2 + c^2}} ; \quad ax + by + cz + d = 0 \tag{4}$$

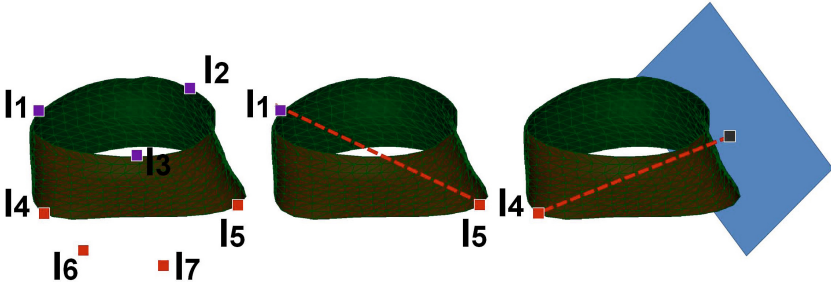


Fig. 3. Geometric features computed based on the sparse landmark representation $L = \{l_1, l_2, l_3, l_4, l_5, l_6, l_7\}$ for the LVOT (left). An example of a distance feature f_{dist} is shown in the middle and a random plane feature f_{rp} on the right.

Distance and random plane features are illustrated in Figure 3. Here, the left diagram shows distance feature ($f_{dist}(l_1, l_5)$), which calculates the Euclidean distance between landmarks l_1 and l_5 associated with the shape. The right diagram shows random plane feature ($f_{rp}(l_4)$), which calculates the shortest distance between landmark l_4 associated with the shape, and a randomly selected plane.

In practice, f_{dist} and f_{rp} were found to be highly complementary when utilized within the ShapeForest, with optimal results produced where given m landmarks, the combined feature set $\{f_{dist} \cup f_{rp}\}$ was computed for each shape s , resulting in a feature set size of $(m^2 - m)/2 + m$ unique features.

3.3 ShapeForest

Spatial features are used in the ShapeForest to obtain a subset of shapes \hat{S} and compute the corresponding constrained SSM. The ShapeForest itself is constructed as a forest of un-pruned decision trees, similar to Breiman’s random forest ensemble classifier [1]. At each tree, each non-leaf node contains a feature $f_\theta \in \{f_{dist} \cup f_{rp}\}$ and threshold value τ , with both leaf and non-leaf nodes further containing a subset of shapes $S_t \in S$. Thus the ShapeForest learns the distance function between the geometric features and the shape variance, clustering shape instances with similar shape characteristics in the leaf nodes. A set of m landmarks $\bar{L} = \{l_1, l_2, \dots, l_m\}$ that is the *GPA aligned mean* of sparse landmarks associated with each shape in S_t is additionally kept at each node in order to facilitate an optional Shape Selection Optimization step as detailed in Section 3.3.

Training. During training, a subset of shapes $S_t = \{s_1, s_2, \dots, s_v\}$ is randomly sampled for each tree, taken from a population of training shapes. Corresponding sets of $LM_t = \{L_1 \cup L_2 \cup \dots \cup L_v\}$ are obtained, where each L_v is the set of sparse landmarks associated with shape s_v in S_t , and spatial features are computed using these landmarks as in Section 3.2. S_t and LM_t are placed at root nodes in

their respective trees, and the following training algorithm is executed for every tree in the ShapeForest:

1. For each feature type $f_\theta \in \{f_{dist} \cup f_{rp}\}$, construct a set of splitting candidates $\phi = (f_\theta, \tau)$, where each τ represents one of a number of threshold values, equally spaced between $\min(f_\theta(LM_t))$ and $\max(f_\theta(LM_t))$.
2. For each ϕ , partition the set of shapes at the current node S_t into left and right subsets:

$$S_l(\phi) = \{s_v | s_v \in S_t \wedge f_\theta(L_v) \leq \tau\} \tag{5}$$

$$S_r(\phi) = \{s_v | s_v \in S_t \wedge f_\theta(L_v) > \tau\} \tag{6}$$

Using GPA, align shapes within subsets S_t , $S_l(\phi)$ and $S_r(\phi)$ to produce aligned sets S_t^a , $S_l^a(\phi)$ and $S_r^a(\phi)$.

3. For each ϕ , compute the information gain $I(S_t, \phi)$ achieved from splitting S_t into $S_l(\phi)$ and $S_r(\phi)$ as:

$$I(S_t, \phi) = \sum_{s \in S_t^a} \log(\delta(s)) - \sum_{i \in \{l,r\}} \left(\sum_{s \in S_i^a(\phi)} \log(\delta(s)) \right) \tag{7}$$

Here, $\delta(s) = s - \bar{s}$ is the deviation of aligned shape s from the mean shape \bar{s} that is calculated from the set of shapes s belongs to (S_t^a , $S_l^a(\phi)$ or $S_r^a(\phi)$), similar to as in [2].

4. Find ϕ^* , the splitting candidate that produces the largest information gain:

$$\phi^* = \underset{\phi}{\operatorname{argmax}} I(S_t, \phi) \tag{8}$$

5. If $I(S_t, \phi^*)$ is greater or equal to a minimum splitting criteria and tree is not at maximum depth, split the node into children, letting $S_t = S_l(\phi^*)$ at the left node and $S_t = S_r(\phi^*)$ at the right node. Update LM_t at each child node accordingly. Finally, save at current node S_t , f_θ and τ from ϕ^* , and \bar{L} as the GPA-aligned mean of landmarks that are associated with each shape in S_t .
6. Repeat steps 1 to 5 of the algorithm at each child node until each tree is fully grown.

cSSM Construction and Tree Aggregation. When presented with a set of sparse landmarks L_u that are associated with a shape, the ShapeForest first computes the feature values for L_u , $f_\theta(L_u) \in \{f_{dist}(L_u) \cup f_{rp}(L_u)\}$. Each individual decision tree is then traversed from their root node through the evaluation of $f_\theta(L_u)$ against τ at each node, branching left or right based on the outcome of this comparison, until a leaf node is reached.

To construct the constrained SSM, the sets of shapes S_t found at the leaf nodes traversed to are returned from every decision tree and aggregated. A shape-frequency histogram is constructed, with the most frequently occurring shapes found across all trees used to set \hat{S} and calculate \hat{s} and $\hat{P}\hat{b}$ of the constrained SSM.

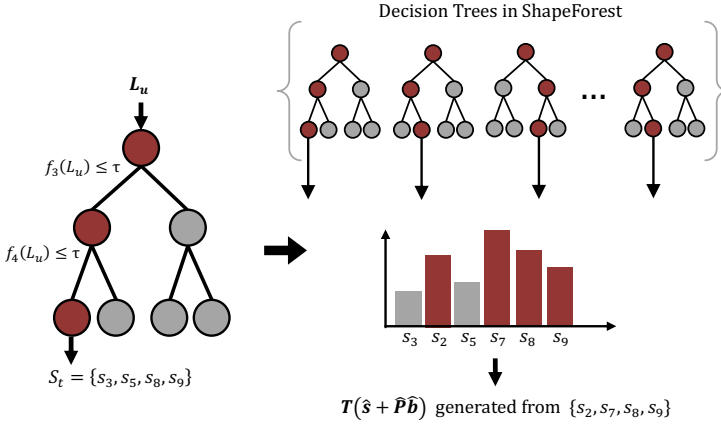


Fig. 4. Example of ShapeForest training and constrained SSM inference

Figure 4 summarizes this process of tree traversal and aggregation described above, showing how a subset of the most frequently occurring shapes (in this case, $\{s_2, s_7, s_8, s_9\}$) can be selected to construct a constrained SSM given the landmarks L .

Shape Selection Optimization (SSO). An optional shape selection optimization step can be applied to all trees in the ShapeForest. This step attempts to utilize the spatial correlation between L_u (the set of sparse landmarks associated with an original shape) and \bar{L} (the set of GPA-aligned mean landmarks stored at a node) to improve node selection after tree traversal, augmenting the set of shapes selection used to construct the constrained SSM.

Selection improvement results in the following algorithm being executed at each tree prior to tree aggregation, starting from selected leaf nodes:

1. At the current node, use GPA to align L_u and \bar{L} and compute corresponding aligned sets L_u^a and \bar{L}^a . Calculate $\delta_c(L_u^a) = \delta(L_u^a) - \delta(\bar{L}^a)$, the deviation between both set of landmarks.
2. Repeat Step 1, but using \bar{L} obtained from the current node's' parent to calculate $\delta_p(L_u^a) = \delta(L_u^a) - \delta(\bar{L}^a)$
3. Calculate $\Delta = \delta_c(L_u^a) - \delta_p(L_u^a)$ as the difference in deviation between L_u and mean landmarks at the current node after alignment, and L_u and mean landmarks at the parent node after alignment. If Δ is greater or equal to a pre-set minimum required value, traverse upwards to the parent node.
4. Repeat Steps 1 to 3 until Δ is less than the pre-set value, or current node is root node.

4 Experiments

To evaluate the ShapeForest, a series of experimental tests were performed. First, in Section 4.2, we evaluated the ability of the ShapeForest to represent complex shape variations. Hereby we compare the ability of the ShapeForest to reconstruct new shapes. Second, in Section 4.3, we integrate the ShapeForest in a standard hierarchical shape segmentation approach (CLM) and compare it with other approaches.

For each experiment, three-fold cross validation experiments were performed. Unless otherwise noted, 100 decision trees of depth 20, each containing 30% of overall training data, were used to construct the ShapeForest, with the top 30% most frequently occurring shapes used to generate each constrained SSM. Where shape selection optimization was used, the Δ threshold was set to 0.

4.1 Datasets

Three data sets were used for our quantitative experiments, each consisting of 3D images containing a specific human anatomy – the left ventricle outflow tract or LVOT (283 shapes), the aortic valve root or AV (633 shapes), and the liver (372 shapes). Each element within each data set contained a three dimensional CT image, a manually annotated shape model s and sparse landmarks set L associated with each shape. Figure 1 illustrates examples of liver shapes (top), AV shapes (middle) and LVOT shapes (bottom).

4.2 Reconstruction Error Evaluation

First we evaluate the ability of the ShapeForest to construct constrained SSMs that can accurately reconstruct new shapes. Given a SSM, the reconstruction error for an unseen shape s_u is computed as follows:

1. Calculate s_u^a as s_u after GPA-alignment with the mean shape defined by the SSM.
2. Project s_u^a onto the subspace defined by the SSM. Calculate s_r as the reconstruction of s_u^a in the SSM subspace using PCA coefficients that correspond to 90% of total variability within the parametric model.
3. Align s_r and s_u using GPA and calculate the reconstruction error as the sum of Euclidean distances between corresponding points in s_r and s_u .

Table 1 summarizes mean reconstruction error results of constrained SSMs produced by the ShapeForest when compared against both the global SSM as well as a multi-modal parametric shape modeling approach, based on methods discussed in [4]. In this alternative approach, the shape model is represented as a mixture of Gaussian parametric models generated by the K-Means clustering of shapes based on their associated landmarks (represented in the table as MixModel2 and MixModel5 depending on the number of mixture components

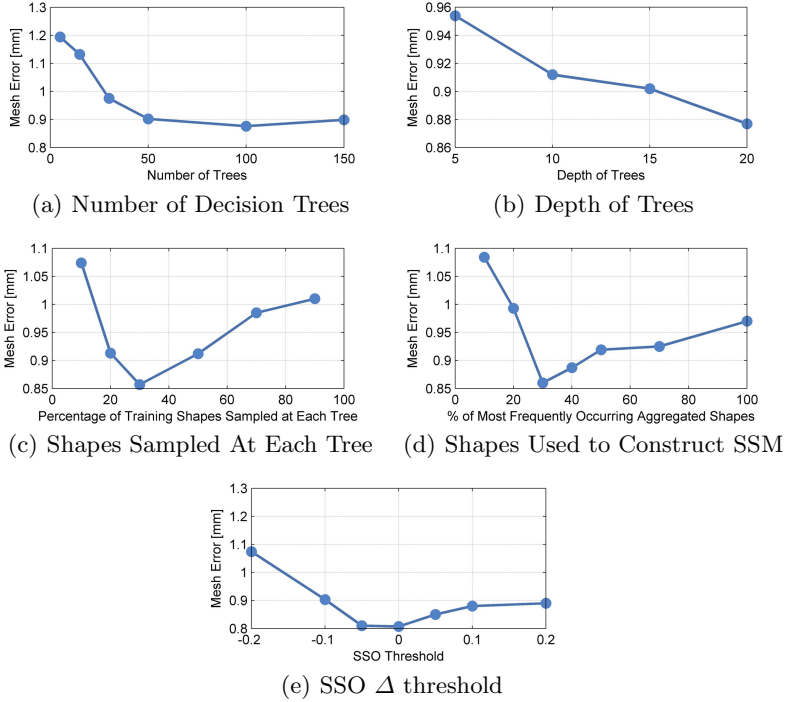


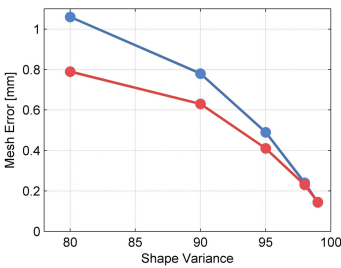
Fig. 5. Effects on segmentation performance of the ShapeForest based on different parameter selection for the LVOT anatomy

used). These experiments show the ShapeForest to consistently be able to reconstruct shapes with a lower error than both the global SSM and the multi-modal parametric shape modeling approach.

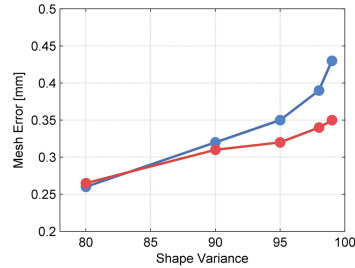
In addition we evaluated the ability of the ShapeForest to approximate the shape manifold with respect to both reconstruction accuracy and specificity for the AV anatomy. The reconstruction accuracy is measured for several instances of shape variability as mentioned in the previous paragraph. The specificity is measuring the ability of the SSM to synthesize realistic shapes, similar to shapes from the ground-truth database. Hereby we synthesize 150 shapes by randomly sampling the shape space coefficients b of the corresponding eigenvectors P within ± 2 times the standard deviation. For the specificity evaluation we measure the distance between the synthetically generated shapes and the closest shape from the ground truth database. Hereby we compared the cSSM+SSO with the gSSM for several percentage of preserved variance in the statistical shape models (see Figure 6). Our proposed method (cSSM+SSO) has better reconstruction accuracy compared to the gSSM with improved or similar specificity for 80%, 90%, 95% and 95% of preserved shape variance. For 99% preserved shape variance the cSSM+SSO reconstruction is in similar range (cSSM+SSO has 1.39% lower reconstruction accuracy) however the specificity improvement using cSSM+SSO is 18.6%.

Table 1. Comparison of reconstruction accuracy of our proposed constrained SSM (and cSSM+SSO) trained with ShapeForest, the global SSM and the Gaussian mixture SSM. The error measurements are computed as the mean Euclidean distance (in millimeters) between the reconstructed shape and the ground-truth test shape. During reconstruction 90% of total variability within the parametric model was used for each SSM approach.

	gSSM	cSSM	cSSM+SSO	MixModel2	MixModel5
LVOT Dataset Error	0.84	0.75	0.71	0.81	0.80
AV Dataset Error	0.78	0.66	0.63	0.71	0.74
Liver Dataset Error	3.27	3.15	3.11	3.32	3.46



(a) Reconstruction accuracy experiment



(b) Specificity experiment

Fig. 6. Reconstruction experiment for the AV anatomy showing the reconstruction accuracy and specificity for both gSSM (blue) and our proposed method, the cSSM+SSO (red)

4.3 Shape Segmentation Evaluation

To examine the performance of the constrained SSMs produced by the ShapeForest, we integrated the cSSM into a hierarchical CLM workflow to segment surface shapes from 3D CT images, as illustrated in Figure 7. We apply the same workflow to segment the AV, LVOT and the liver.

Based on a large set of annotations, we train a position detector for each of the landmarks L independently. We use a discriminative learning approach (Probabilistic Boosting Tree) trained with Haar-like features [20] to delineate the points in the images. Based on the landmarks the constrained SSM is constructed using the ShapeForest. For the Gaussian mixture model the cluster (representing an unique Gaussian parametric model) was selected based on the Euclidean distance of the aligned detected landmarks L to the cluster mean landmark model.

Based on the detected landmarks and one of the shape space representations an initial shape model is fitted to the image data. Hereby we use the Powell optimization consecutively to estimate the coefficients for the first 5 largest

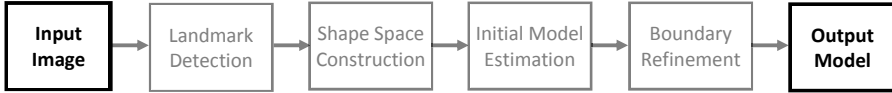


Fig. 7. Hierarchical shape segmentation workflow

eigenvectors. Starting with the largest eigenvector the best value is found to match accurately the landmarks L in the images. Hereby no image information is used to fit the model. Finally an iterative CLM method is applied to locally refine the initialized shape. The feature response images are computed with the probabilistic boosting tree and steerable features. Iteratively the current shape is adjusted along its local curvature based on the feature response images. Later the updated shape is projected along the basis vectors of each corresponding shape space, covering 99% of the shape variation in the training data.

Using this workflow, tests were performed to quantitatively evaluate the accuracy of shape models produced with constrained SSM's generated by the ShapeForest against shape models produced with the global SSM, and the multi-modal parametric shape modeling approach described in Section 4.2. Table 2 summarizes the results of these experiments, showing CLM with constrained SSMs produced by the ShapeForest to consistently outperform both the global SSM and the multi-modal parametric shape model in each data set. For the LVOT and AV data set the mixture model improved the accuracy compared to the global SSM. However it failed for the liver data as the parametric model assignment based on the Euclidean distance between landmarks was inadequate in many cases. Execution overhead as a result of utilizing constrained SSMs was minimal (< 0.02 second per segmented shape) across all data sets.

Figure 8 shows examples of shapes which are used to construct the global SSM (left column) and the constrained SSM (middle and left column) for the LVOT, AV and the liver. It demonstrates how the ShapeForest clusters shapes with similar geometric properties, preserving local detail in the selected shape population. Parametric models inferred by the ShapeForest approximate more accurately the variation for a new instance, compared to parametric models build from the complete shape population.

4.4 Effect of ShapeForest Parameters

To examine the influence of the ShapeForest parameter selection used for constrained SSM construction, experiments were performed within the CLM segmentation framework for the LVOT data set. Using the parameters in the experiments as described in Section 4.3 with no SSO applied as a baseline, each of the five primary parameters that control constrained SSM construction in the ShapeForest were varied. Results of these experiments are summarized in Figure 5.

Table 2. Comparison of segmentation accuracy of our proposed constrained SSM (and cSSM+SSO) trained with ShapeForest, the global SSM and the Gaussian mixture SSM. The error measurements are computed as the mean Euclidean distance (in millimeters) between the segmented shape and the manually annotated ground-truth shape. 99% of total variability within the parametric model was used for each SSM approach.

Method	Dataset	Shapes	Landm	Error in Initial Model				Error in Final Model			
				Mean	STD	Median	Improv	Mean	STD	Median	Improv
gSSM	LVOT	283	7	1.65	0.61	1.51	-	1.02	0.41	0.89	-
cSSM	LVOT	283	7	1.32	0.59	1.26	+19.69%	0.84	0.33	0.79	+17.64%
cSSM+SSO	LVOT	283	7	1.29	0.56	1.23	+21.81%	0.81	0.25	0.78	+20.58%
MixModel2	LVOT	283	7	1.51	0.53	1.46	+8.4%	0.88	0.33	0.80	+13.7%
MixModel5	LVOT	283	7	1.50	0.51	1.38	+9.1%	0.87	0.31	0.76	+14.7%
MixModel10	LVOT	283	7	2.73	1.19	2.42	-65.5%	1.82	1.48	1.32	-78.4%
gSSM	AV	633	13	0.99	0.50	0.85	-	0.83	0.20	0.82	-
cSSM	AV	633	13	0.80	0.38	0.79	+19.19%	0.69	0.29	0.63	+16.87%
cSSM+SSO	AV	633	13	0.78	0.32	0.62	+21.21%	0.66	0.19	0.61	+20.48%
MixModel2	AV	633	13	0.85	0.15	0.84	+14.1%	0.76	0.21	0.70	+8.4%
MixModel5	AV	633	13	0.93	0.19	0.91	+10.0%	0.78	0.24	0.75	+6.0%
MixModel10	AV	633	13	1.25	0.64	1.08	-26.3%	0.88	0.20	0.87	-6.0%
gSSM	Liver	372	20	3.99	0.95	3.82	-	2.85	0.92	2.68	-
cSSM	Liver	372	20	3.42	1.19	3.37	+14.28%	2.29	0.90	2.16	+19.64%
cSSM+SSO	Liver	372	20	3.34	1.29	2.86	+16.29%	2.20	0.76	2.16	+22.8%
MixModel2	Liver	372	20	4.05	0.99	3.89	-1.5%	2.88	1.23	2.67	-1.0%
MixModel5	Liver	372	20	4.44	1.28	4.37	-11.2%	3.54	1.54	3.31	-24.2%
MixModel10	Liver	372	20	4.82	1.26	4.94	-20.1%	4.13	1.41	4.16	-44.9%

Number of Decision Trees: The effect of increasing the number of decision trees used in the ShapeForest has on segmentation accuracy is shown in Figure 5(a). Here, increasing the number of trees decreased the mean error progressively, until reaching a plateau once the number of trees had been increased to 100. Increasing the number of trees beyond this point had negligible impact on segmentation accuracy while considerably increasing the memory requirements of the ShapeForest.

Depth of Decision Trees: The effect of increasing the depth of each decision trees is shown in Figure 5(b). A progressive decrease in mean error was noted as depth was increased, with optimal results found when trees were left to grow completely unbounded. Utilizing deeper decision trees increases the memory requirements of the ShapeForest considerably, although contributes only minimal additional overhead to execution time.

Sampled Data at Each Decision Tree: The degree that the amount training data used to construct each decision tree can affect mean error is shown in Figure 5(c), with optimal values observed when 30% of all training data is randomly sampled at each tree. Sampling a small amount of data resulted in trees that shared an insufficient amount of shapes to accurately obtain a set of frequently occurring shapes, while sampling a large amount of data resulted in individual trees being too similar to each other, minimizing the benefit gained from random sampling.

Percentage of Shapes used in SSM Construction: Figure 5(d) shows the effects of increasing the percentage of frequently occurring shapes used to construct the constrained SSM, based on the shape frequency histogram computed based on the aggregation of the results of all decision trees in the ShapeForest. Taking the top 30% most frequently occurring shapes produced the lowest error.

Shape Selection Optimization: Usage of the optimal SSO step was found to be able to improve the accuracy of shape detection, as shown in Figure 5(e), with optimal values found in the range of -0.05 to $+0.05$. Progressively lower values caused each tree to move back to their root node, with progressively higher values preventing any traversal upwards from the selected leaf node.

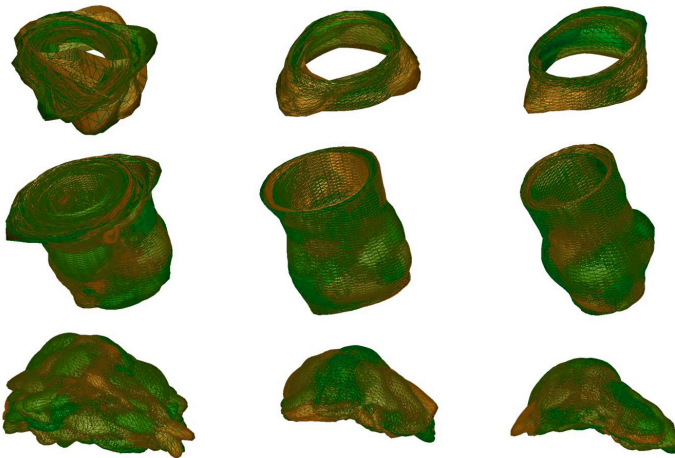


Fig. 8. Aligned Shapes using GPA for the global SSM (left) and inferred shapes with ShapeForest used to compute the constrained (mid, right) SSMs for the LVOT (top), AV (mid) and Liver (bottom) anatomy

5 Conclusion

In this paper we have presented ShapeForest, a novel shape model representation trained with decision trees and geometric features. It enables modeling of complex shape variation, preserves local detail in the training data and incorporates lower-level information in form of sparse landmarks during shape space learning and shape inference. Quantitative experiments on three separate medical data sets have shown that the usage of the constrained statistical shape model produced by the ShapeForest leads to substantial improvement in shape modeling and shape segmentation when incorporated in a Constrained Local Model framework.

References

1. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
2. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)
3. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 681–685 (2001)
4. Cootes, T.F., Taylor, C.J.: A mixture model for representing shape variation. *Image and Vision Computing* 17(8), 567–573 (1999)
5. Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K.: Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical Image Analysis* (2013)
6. Cristinacce, D., Cootes, T.: Automatic feature localisation with constrained local models. *Pattern Recognition* 41(10), 3054–3067 (2008)
7. Davatzikos, C., Tao, X., Shen, D.: Hierarchical active shape models, using the wavelet transform. *IEEE Transactions on Medical Imaging* 22(3), 414–423 (2003)
8. Etyngier, P., Segonne, F., Keriven, R.: Shape priors using manifold learning techniques. In: *IEEE 11th International Conference on Computer Vision, ICCV 2007*, pp. 1–8. IEEE (2007)
9. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. *International Journal of Computer Vision* 61(1), 55–79 (2005)
10. Gall, J., Lempitsky, V.: Class-specific hough forests for object detection. In: *Decision Forests for Computer Vision and Medical Image Analysis*, pp. 143–157. Springer (2013)
11. Georgescu, B., Zhou, X.S., Comaniciu, D., Gupta, A.: Database-guided segmentation of anatomical structures with complex appearance. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. 2, pp. 429–436. IEEE (2005)
12. Grbic, S., Ionasesc, R., Vitanovski, D., Voigt, I., Wang, Y., Georgescu, B., Comaniciu, D.: Complete valvular heart apparatus model from 4d cardiac ct. *Medical Image Analysis* 16(5), 1003–1014 (2012)
13. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
14. Rodola, E., Buló, S.R., Windheuser, T., Vestner, M., Cremers, D.: Dense non-rigid shape correspondence using random forests (2014)
15. Saragih, J., Goecke, R.: A nonlinear discriminative approach to aam fitting. In: *IEEE 11th International Conference on Computer Vision, ICCV 2007*, pp. 1–8. IEEE (2007)
16. Saragih, J.M., Lucey, S., Cohn, J.F.: Deformable model fitting with a mixture of local experts. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 2248–2255. IEEE (2009)
17. Sjostrand, K., Rostrup, E., Ryberg, C., Larsen, R., Studholme, C., Baezner, H., Ferro, J., Fazekas, F., Pantoni, L., Inzitari, D., et al.: Sparse decomposition and modeling of anatomical shape variation. *IEEE Transactions on Medical Imaging* 26(12), 1625–1635 (2007)
18. Yang, L., Georgescu, B., Zheng, Y., Meer, P., Comaniciu, D.: 3d ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008*, pp. 1–8. IEEE (2008)

19. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Sparse shape composition: A new framework for shape prior modeling. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1025–1032. IEEE (2011)
20. Zheng, Y., Georgescu, B., Ling, H., Zhou, S.K., Scheuring, M., Comaniciu, D.: Constrained marginal space learning for efficient 3d anatomical structure detection in medical images. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 194–201. IEEE (2009)