Active Shape Models Unleashed

Matthias Kirschner and Stefan Wesarg

Technische Universität Darmstadt, Interactive Graphics Systems Group
Fraunhoferstrasse 5, 64283 Darmstadt, Germany

ABSTRACT

Active Shape Models (ASMs) are a popular family of segmentation algorithms which combine local appearance models for boundary detection with a statistical shape model (SSM). They are especially popular in medical imaging due to their ability for fast and accurate segmentation of anatomical structures even in large and noisy 3D images. A well-known limitation of ASMs is that the shape constraints are over-restrictive, because the segmentations are bounded by the Principal Component Analysis (PCA) subspace learned from the training data. To overcome this limitation, we propose a new energy minimization approach which combines an external image energy with an internal shape model energy. Our shape energy uses the Distance From Feature Space (DFFS) concept to allow deviations from the PCA subspace in a theoretically sound and computationally fast way. In contrast to previous approaches, our model does not rely on post-processing with constrained free-form deformation or additional complex local energy models. In addition to the energy minimization approach, we propose a new method for liver detection, a new method for initializing an SSM and an improved k-Nearest Neighbour (kNN)-classifier for boundary detection. Our ASM is evaluated with leave-one-out tests on a data set with 34 tomographic CT scans of the liver and is compared to an ASM with standard shape constraints. The quantitative results of our experiments show that we achieve higher segmentation accuracy with our energy minimization approach than with standard shape constraints.

Keywords: Active Shape Model, segmentation, shape constraints, statistical shape model.

1. INTRODUCTION

Active Shape Models (ASMs) have been successfully used for fast and accurate segmentation of anatomic structures in 3D medical images. An ASM uses a statistical shape model (SSM) for constraining the detected organ boundary to 'plausible' shapes, that means shapes similar to those in the model's training data set. However, due to the large variability of anatomic structures and the limited number of available training examples, the SSM can not completely capture all possible variations of an object class. Moreover, pathologies may cause large deviations from the natural organ shape. The standard approach of ASMs for constraining shapes[1] does not account for the limited generalization ability of the SSM. Deformed shapes are constrained by projecting them to the subspace spanned by the principal eigenvectors of the data covariance matrix, and by imposing bounds on the resulting vector. This approach is over-restrictive, because the subspace typically does not contain every possible shape variation of the structure of interest (see also Figure 1).

Several strategies have been proposed to overcome the limitations of the standard ASM algorithm. Many ASM variants[2–4] refine the segmentations by relaxing shape constraints, thus achieving higher segmentation accuracy than the classical ASM. On the downside, these extensions come at the cost of increased model complexity. For example, Heimann et al.[3] define additional tension and rigidity forces to allow shapes to leave the SSM's shape space. These additional forces are controlled by several ad-hoc parameters, for which there is no obvious a priori choice. In general, time consuming experiments are necessary to tune these parameters to a specific application.

An alternative and complementary approach for relaxing shape constraints is to artificially increase the size of the training data set.[5] The general idea is to synthesize new training shapes by manipulating and combining existing ones. Koikkalainen et al.[5] have shown that the additional flexibility introduced by these techniques can indeed lead to more accurate segmentation. However, additional models must be designed for synthesizing new shapes, and these models are often based on arbitrary assumptions. Similarly to ASMs with relaxed

Send correspondence to M.K.: matthias.kirschner@gris.tu-darmstadt.de; phone +49 6151 155624

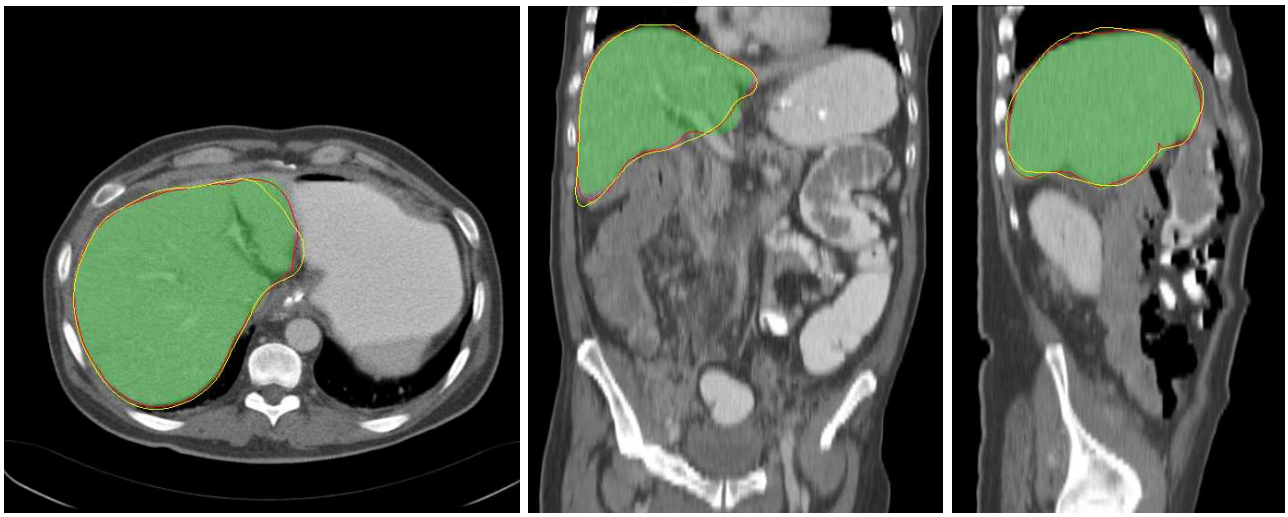

Figure 1. Exemplary segmentation results on one of our test images (from left to right: transversal, coronal and sagittal view). Green area: Ground truth segmentation. Yellow Curve: Outline of the segmentation computed by ASM with standard shape constraints.[1] Red Curve: Outline of the segmentation computed using our new shape energy ($\beta = 5$). The segmentation computed with our new approach delineates the boundary better than the standard approach, because the shape energy is less restrictive than the standard shape constraints.

shape constraints, the generation of artificial shapes is steered by ad-hoc parameters, which must be tuned to a particular application. Moreover, a standard SSM learned from an enlarged training set is still restricted to a linear subspace of the original input data space in which the training data lives.

Instead of working in a linear subspace, Cremers et al.[6] constrain shapes by minimizing an energy in the original data space. Their segmentation algorithm balances texture information and shape energy, where the latter term is the Mahalanobis distance to the mean shape. This implies the assumption that the shapes are Gaussian distributed, similar to the original SSM. While this approach can be easily used for 2D image segmentation, the huge dimensionality of 3D SSMs quickly becomes a problem. For a 3D SSM with 2562 landmarks, which is the size we use in our experiments, we would have to work with a matrix with more than 59 million entries during optimization, because the approach of Cremers et al.[6] does not reduce the model's dimensionality.

In this paper, we propose a new ASM algorithm for 3D medical images. The key contribution is a new shape energy term, which allows to constrain shapes in the original data space, while it uses at the same time Principal Component Analysis (PCA) to reduce the model's dimensionality. Our approach has the following properties:

1. Shapes can leave the linear subspace spanned by the first principal components, which is not possible in the classical ASM approach[1] .
2. Our shape energy term is based on a solid theoretical foundation. The only assumptions we make is that the shapes are Gaussian distributed. This stands in contrast to hybrid approaches,[2–4] which introduce additional assumptions and thus increase the model complexity.
3. Our shape energy term does not require additional parameters, whereas hybrid approaches typically rely on ad-hoc parameters.
4. Our model uses PCA to reduce the model's dimensionality. Thus, energy minimization in the original data space becomes feasible even for high dimensional 3D SSMs.

Our new energy minimization approach is incorporated into an ASM algorithm for liver segmentation. We evaluate it on a set of 34 contrast enhanced CT scans of the liver with leave-one-out tests, and compare it with standard shape constraints. With regard to the ASM algorithm we use for evaluation, we make the following additional contributions:

- We present a new approach for automatic detection of the liver in contrast-enhanced CT images.
- We present a modification of the Iterative Closest Point (ICP) algorithm,[7] which can register an SSM to a point set. The original ICP is only capable of registering a fixed mesh (e.g. the mean shape) to a point set. The modified ICP algorithm is used for automatically initializing pose and shape parameters of the SSM.
- We propose an improvement of kNN-classifier-based appearance models. Our classifier combines oriented features with foreground/background classification. Moreover, the classifier is built using fuzzy class labels, thus accounting for uncertainty at the boundary of the ground-truth segmentation.

2. METHODS

2.1 Statistical Shape Models and Active Shape Model Segmentation

ASM segmentation[1] uses a landmark-based shape representation. Each shape is represented by a shape vector $\boldsymbol{x} = \{x_1, y_1, z_1, \ldots, x_N, y_N, z_N\} \in \mathbb{R}^{3N}$, that is constructed by concatenating the N points $\boldsymbol{x}(j) = \{x_j, y_j, z_j\} \in \mathbb{R}^3$, $j \in \{1, \ldots, N\}$, which are located on the shape's boundary. Points on different shapes, but with the same index, correspond to each other. An SSM is built by applying PCA to a set of S training shape vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_S$. The PCA yields the eigenvectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{3N}$ and corresponding eigenvalues $\lambda_1 \geq \ldots \geq \lambda_{3N}$ of the covariance matrix $\boldsymbol{\Sigma} = \frac{1}{S-1}\sum_{i=1}^{S}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T$, where $\bar{\boldsymbol{x}} = \frac{1}{S}\sum_{i=1}^{S}\boldsymbol{x}_i$ is the mean shape. By discarding eigenvectors with small eigenvalue, the dimensionality of the model can be reduced. In this work, we discard all eigenvalues with index $i > t$, were $t = \min\{t' | \sum_{i=1}^{t'} \lambda_i / \sum_{i=1}^{3N} \lambda_i > 98\%\}$. By $\boldsymbol{P} = (\boldsymbol{p}_1 | \ldots | \boldsymbol{p}_t)$ we denote the $3N \times t$-matrix of the retained eigenvectors.

The ASM segmentation algorithm starts with placing the mean shape of the SSM into the image. In an iterative procedure, landmarks are displaced in the matching step to image features in their vicinity, which results in a deformed shape $\hat{\boldsymbol{x}}$. The quality of image features is assessed by a local appearance model. The deformed shape $\hat{\boldsymbol{x}}$ is then constrained to a shape which is 'plausible' with respect to the learned SSM. The standard approach is to project $\hat{\boldsymbol{x}}$ to the subspace F spanned by the principal axes using the equation $\boldsymbol{b} = \boldsymbol{P}^T(\hat{\boldsymbol{x}} - \bar{\boldsymbol{x}})$. Then, for each i, the i-th component of the subspace vector $\boldsymbol{b} \in \mathbb{R}^t$ is clamped to the interval $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$. From $\boldsymbol{b}$, a plausible shape $\boldsymbol{x}$ is computed using the formula $\boldsymbol{x} = \hat{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{b}$.

2.2 Constraining Shapes with Energy Minimization

We propose to constrain deformed shapes during ASM segmentation with an energy minimization approach. After each matching step of the ASM algorithm, the new plausible shape $\boldsymbol{x}^*$ is determined by minimizing the energy function

$$E(\boldsymbol{x}) = E_{\text{image}}(\boldsymbol{x}) + E_{\text{shape}}(\boldsymbol{x}). \tag{1}$$

Analogous to various other approaches in image segmentation, our energy function consists of an external energy $E_{\text{image}}(\boldsymbol{x})$ that draws the contour of the model towards the optimal image features found during matching, and an internal energy $E_{\text{shape}}(\boldsymbol{x})$ that penalizes deviations from the shape model.

2.2.1 Shape Energy

If we assume – as we do in the classical shape model – that the shapes are multivariate normally distributed, then a suitable choice for the $E_{\text{shape}}(\boldsymbol{x})$ would be the negative log-likelihood of the probability distribution

$$\Pr(x) = \frac{1}{(2\pi)^{\frac{3N}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}})\right) \tag{2}$$

where mean $\bar{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{\Sigma}$ are estimated from the data. However, two practical considerations must be taken into account: Firstly, the dimension $3N$ is typically in the order of several thousands. Handling a squared matrix of this size would slow down the optimization time considerably. Secondly, since we typically have far less training examples than dimensions, i.e. $S \ll 3N$, the estimated covariance matrix is singular, because the rank of $\boldsymbol{\Sigma}$ can not exceed $S - 1$.

We therefore propose to work with the much smaller matrix P of retained eigenvectors of $\mathbf{\Sigma}$, as in the classical shape model, but without restricting the model to the linear subspace $F \subset \mathbb{R}^{3N}$ spanned by these eigenvectors. We adopt the idea of penalizing the *Distance From Feature Space* (DFFS), which was proposed by Moghaddam and Pentland.[8] The DFFS is simply the cost of the projection of a vector $\boldsymbol{x}$ to F, i.e. $\|\boldsymbol{r}\|^2 = \|\boldsymbol{x} - (\bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{b})\|^2 = \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2 - \|\boldsymbol{b}\|^2$, where $\boldsymbol{r}$ denotes the residual vector of the projection. Note that $\|\boldsymbol{r}\|^2$ can be completely expressed in terms of $\boldsymbol{x}$, $\bar{\boldsymbol{x}}$ and $\boldsymbol{P}$. Moghaddam and Pentland propose to use the DFFS in order to approximate the probability function $\Pr(\boldsymbol{x})$ with the product of two marginal and independent Gaussians:

$$\hat{\Pr}(\boldsymbol{x}) = \Pr_F(\boldsymbol{x}) \cdot \hat{\Pr}_{F^\perp}(\boldsymbol{x}) = \left[\frac{1}{(2\pi)^{\frac{t}{2}} \prod_{i=1}^{t} \lambda^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{t} \frac{\boldsymbol{b}_i^2}{\lambda_i}\right)\right] \cdot \left[\frac{1}{(2\pi\rho)^{\frac{3N-t}{2}}} \exp\left(-\frac{1}{2\rho}\sum_{i=t+1}^{3N} \boldsymbol{r}_i^2\right)\right]. \quad (3)$$

Here, $F^\perp$ denotes the orthogonal complement of F. The parameter ρ controls how strong the DFFS is penalized. It can be shown[8] that $\rho^* = \frac{1}{3N-t}\sum_{i=t+1}^{3N} \lambda_i$ is the optimal choice for this parameter with respect to the Kullback-Leibler divergence $J(\rho) = E\left[\log \frac{\Pr(\boldsymbol{x})}{\hat{\Pr}(\boldsymbol{x})}\right]$. For the shape energy, we therefore use the negative log-likelihood (up to a constant term) of $\hat{\Pr}(\boldsymbol{x})$:

$$E_{\text{shape}}(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{t} \frac{\boldsymbol{b}_i^2}{\lambda_i} + \frac{1}{2\rho^*}\|\boldsymbol{r}\|^2. \quad (4)$$

In contrast to Cremers et al.,[6] we reduce the model's dimensionality by discarding eigenvectors with small eigenvalue. A smaller dimensionality has the advantage that we need less time for optimization and save memory. Moreover, eigenmodes with small eigenvalue may not be genuine shape variations, but noise. Sources of noise can be, for example, inaccuracy of manual expert segmentations, or the process that computes training shape vectors from these segmentations.

2.2.2 Image Energy

The image energy is used to penalize deviations from the deformed shape $\hat{\boldsymbol{x}}$. We use

$$E_{\text{image}}(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{\|\boldsymbol{x}(i) - \hat{\boldsymbol{x}}(i)\|^2}{2\sigma_i^2}. \quad (5)$$

The parameters σ_i are chosen in dependence of the weights $w_i \in [0, 1]$ of the texture feature, which were determined by the local appearance model (see section 2.4). The idea is to allow greater deviation (higher σ_i) from the proposed landmark position if w_i is small. We set $\sigma_i = \beta^{-1}\exp(1 - w_i)$, where β is a free parameter. Similar to Heimann et al.,[3] we select the texture features that define the external energy with the graph theoretical *optimal surface detection* algorithm.[9] This removes outlying texture features and leads to a smoother deformed shape $\hat{\boldsymbol{x}}$.

2.2.3 Optimization

The domain of the energy function $E(\boldsymbol{x})$ is the space $\mathbb{R}^{3N}$ (in our case $3N = 7686$). In order to avoid long optimization times, we must choose an optimizer which runs quickly even in such a high dimensional input space. We therefore selected the limited memory Broyden-Fletcher-Goldfarb-Shannon method[10] (L-BFGS), which performs well in our experiments (see section 4). We use the implementation of the L-BFGS method in the vxl library (`http://vxl.sourceforge.net/`). Note that the gradient $\nabla E(\boldsymbol{x})$ of the energy function, which is required by the L-BFGS method, can be computed analytically.

2.3 Automatic Shape Model Initialization

Like all ASM algorithms, our algorithm is a local search algorithm. This means that a good initial placement of the model into the image is necessary for automatic segmentation. Since we evaluate our algorithm on a set of contrast enhanced CT scans of the liver, we need to detect the liver in the image in order to determine the initial model position. For liver detection, we developed an algorithm which combines ideas from Kainmüller et al.[2] and Ruskó et al.[11] After this liver detection phase, we register our SSM with points on the liver contour using

a new modified version of the ICP algorithm, which does not only estimate pose parameters, but also shape parameters. While the detection is tailored to the application of liver segmentation, the registration is a general method for placing a 3D SSM into an image.

We start with estimating the intensity range of the liver tissue. First, we fit a Gaussian mixture model to the image histogram,[2] restricted to the interval of [-50, 300] HU. We select the highest peak in the mixture model with HU value > 80 to be the liver tissue,[11] which is characterized by its mean μ_l and standard deviation σ_l. We use thresholding in order to map all voxels within the interval $[\mu_l - 2\sigma, \mu_l + 2\sigma]$ to 1 and all other voxels to 0. We reduce the noise in this binary feature image with a pipeline of image filters, which includes morphological filters like opening and hole filling as well as a connected component analysis filter. We then estimate the liver's bounding box from the feature image by searching for the subimage with the most voxels that are 1, because the liver is the largest organ. The size of this subimage can be estimated from the known liver bounding boxes in the training data.

In the next step, we identify voxels that lie on the liver boundary. We apply a contour filter to the binary feature image and a gradient magnitude filter followed by a thresholding to the original image. These two binary images are then combined with an AND operation. The complete filter pipeline is illustrated in Figure 2. From the image computed by this pipeline, we sample a subset of 8000 voxels in order to build a point set $\mathcal{M}$, and then register the SSM to these points using an extended version of the ICP algorithm. The classical ICP algorithm[7] rigidly registers a point set $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_L\}$ with another point set $\mathcal{M} = \{\boldsymbol{m}_1, ..., \boldsymbol{m}_K\}$. In each iteration i of the classical approach, a correspondence function $C^i : \mathcal{V} \mapsto \mathcal{M}$ between points from $\mathcal{V}$ and $\mathcal{M}$ is estimated, which is defined by the formula $C^i(\boldsymbol{v}) = \text{argmin}_{\boldsymbol{m} \in \mathcal{M}} \|(\boldsymbol{R}^i \boldsymbol{v} + \boldsymbol{t}^i) - \boldsymbol{m}\|$. Here, $\boldsymbol{R}^i$ and $\boldsymbol{t}^i$ denote the estimates of the rotation matrix and translation vector in iteration i. On the basis of C^i, the updated estimates $\boldsymbol{R}^{i+1}$ and $\boldsymbol{t}^{i+1}$ are derived using the method of Horn.[12]

Our extension of the ICP algorithm registers a complete SSM with $\mathcal{M}$, instead of a static point set. In this algorithm, a vector $\boldsymbol{x}^i$ represents the point set $\mathcal{V}$, which means that we have $\boldsymbol{v}_j = \boldsymbol{x}^i(j)$ in iteration i. We initialize $\boldsymbol{x}^1$ with the mean shape $\bar{\boldsymbol{x}}$. In each iteration i, after we estimated the new pose parameters $\boldsymbol{R}^{i+1}$ and $\boldsymbol{t}^{i+1}$, we deform $\boldsymbol{x}^i$ to $\hat{\boldsymbol{x}}^i$ using the function C^i. We use the pose parameters to project the deformed vector $\hat{\boldsymbol{x}}^i$ into the shape model's coordinate system, where we constrain it to a plausible shape using the formula $\boldsymbol{x}^{i+1} = \bar{\boldsymbol{x}} + \boldsymbol{P}(\boldsymbol{P}^T(\hat{\boldsymbol{x}}^i - \bar{\boldsymbol{x}}))$. This plausible shape is then reprojected to world space coordinates, and the next iteration begins. In contrast to the classical ICP algorithm, we also estimate a scale factor using Horns method.[12] By using a k-d-tree to determine C^i in each iteration, the described algorithm is very efficient.

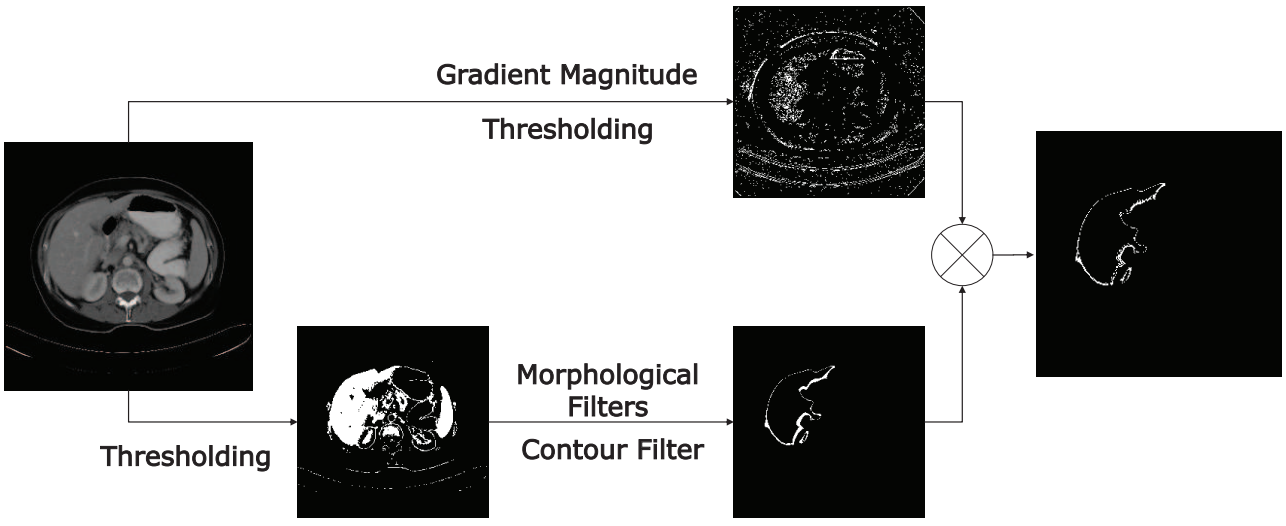

Figure 2. A pipeline of image filters is applied to the original image (leftmost image) in order to find an approximate liver boundary (rightmost image). From the resulting image, we sample a subset of voxels, and register the SSM against this point set in order to initialize the ASM algorithm.

The ICP algorithm must be initialized with pose parameters $s^1, \boldsymbol{R}^1, \boldsymbol{t}^1$ for the first iteration. Estimates of the scale factor s^1 and the translation $\boldsymbol{t}^1$ are derived from the detected liver bounding box. Since the orientation of the liver is similar in most images due to the CT acquisition protocol, taking $\boldsymbol{R}^1$ to be the identity matrix usually performs well. This is because the identity matrix corresponds to the 'mean' liver pose of the training data. However, like the original ICP algorithm, our modified ICP algorithm can get trapped in local optima. Therefore, we additionally increase the robustness by running the algorithm 15 times, using three different scaling factors and five different rotation matrices. The rotation matrices we use describe rotations around the y-axis by no more than 45°. Each of the 15 registration results is a possible candidate for the initial liver shape. We choose the candidate which minimizes

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \|\boldsymbol{x}(i) - C(\boldsymbol{x}(i))\|^2 + \sum_{i=1}^{t} \frac{\boldsymbol{b}_i(\boldsymbol{x})^2}{\lambda_i} \tag{6}$$

where $\boldsymbol{x}$ is a shape with N points $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)$, $C(\boldsymbol{x})$ is the final correspondence relation between $\mathcal{V}$ and $\mathcal{M}$, and $\boldsymbol{b}(x)$ are the shape parameters of $\boldsymbol{x}$. Note that the first term measures how good $\boldsymbol{x}$ fits to $\mathcal{M}$, whereas the second term measures the likelihood of the shape.

2.4 Local Appearance Models

We use kNN-classifiers to detect the organ boundaries. kNN-classifiers were first applied to ASMs by van Ginneken et al.[13] In the original work, image features are sampled in a rectangular grid around each landmark and then labeled with the class labels *background* and *foreground*. In contrast to classical Gaussian intensity profiles, the features are not oriented along the surface normal. During segmentation, several features are sampled on a profile in the direction of the surface normal and then classified with a kNN-classifier learned during training. The landmark position is moved to the position on the profile which best separates between foreground and background. Heimann et al.[3] later adopted the idea of kNN-classifiers, but in contrast to van Ginneken et al., they use features oriented along the surface normal. Moreover, they classify features in *boundary* and *non-boundary* features, instead of distinguishing between foreground and background.

In our opinion, oriented features are more meaningful, but the classes foreground and background are easier separable than boundary and non-boundary. Therefore, we use oriented features, but foreground and background as class labels. Additionally, we propose to use fuzzy class labels, because it is difficult to classify features close to the surface boundary. During training, we sample for each landmark a set of image features by placing a profile at the landmark position and shifting it inside and outside the object. The profile is oriented along the surface normal. Each sampled feature consists of three intensity values. The coordinates of a feature's center are mapped to the reference segmentation in order to determine the class label. Because we use linear interpolation to read the value on the reference segmentation image, we get fuzzy class labels on points close to the surface boundary.

During segmentation, we consider profiles $\boldsymbol{y} = (\boldsymbol{y}_{-k}, \ldots, \boldsymbol{y}_k)$ of length (2k+1), where the $\boldsymbol{y}_i \in \mathbb{R}^3$ are points on the profile. The index of these points runs from the inside to the outside of the segmentation boundary. For each point $\boldsymbol{y}_i$ on the profile, an image feature is sampled. We compute the probability $c(\boldsymbol{y}_i) \in [0, 1]$ that the point $\boldsymbol{y}_i$ is a foreground point, i.e. that it is inside the object, using the kNN-classifier. The fitness of the profile is then $f(\boldsymbol{y}) = 1.0 - |0.5 - c(\boldsymbol{y}_0)| + \sum_{i=-k}^{-1} c(\boldsymbol{y}_i) + \sum_{i=1}^{k} (1 - c(\boldsymbol{y}_i))$. The equation follows the formula of van Ginneken et al.,[13] with the exception of the first term, which rewards indecisiveness of the classifier at the boundary between foreground and background. The displaced landmark is set to the point $\boldsymbol{y}_0^*$ of the profile $\boldsymbol{y}^*$ which minimizes $f(\boldsymbol{y})$. To determine landmark weights w_i for $E_{\text{image}}(\boldsymbol{x})$, we normalize the fitness values sampled along a single landmark such that they sum up to 1, and use the normalized fitness of the winner profile as weight.

We apply our appearance model to contrast enhanced CT scans of the liver. Because of the contrast agent, the liver tissue is not always in the same HU interval in different images. This can degrade our appearance model, because it uses intensity based image features. We therefore normalize the image features by subtracting the mean liver intensity. For the training data, we can use the true mean liver intensity, and during segmentation, we use the approximated mean liver intensity which we estimated for liver detection (see section 2.3).

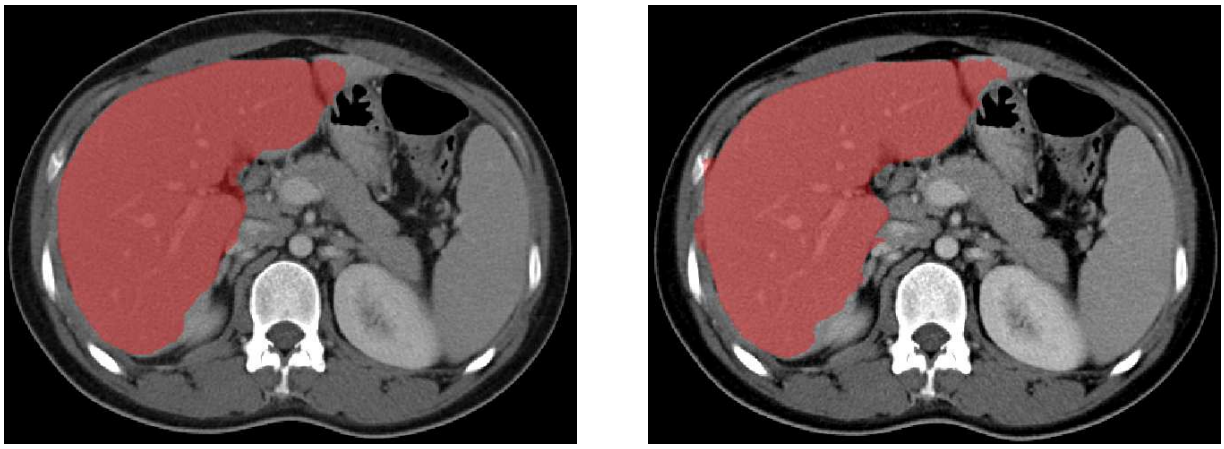

Figure 3. Transversal view on one of our test images. We segmented the image with our ASM algorithm with parameters $\beta = 4$ (left image) and $\beta = 6$ (right image). The segmentation result in the left image is smoother, which leads to higher segmentation accuracy on this image.

3. EXPERIMENTS

Our evaluation has two goals: Firstly, we want to compare our new energy minimization approach against standard shape constraints,[1] which are still being used in many ASM algorithms. Secondly, we want to investigate the influence of the only free parameter, β, which influences the force of the texture energy.

We use a set of 34 contrast enhanced CT scans of the liver for our tests. The spacing of the data sets ranges from 0.55 to 0.8 mm in x– and y–direction, whereas the slice thickness ranges from 1 to 5 mm. For each data set, a reference segmentation is given. From these reference segmentations, meshes were extracted and correspondence was established using an automatic algorithm,[14] resulting in training shapes with 2562 landmarks each. The landmark positions were used to extract training image features from each image for the kNN-classifier.

We evaluate our ASM with energy minimization in leave-one-out tests, i.e. for segmenting the i-th data set, both shape and appearance model are trained without using any data from the i-th shape or image. As a preprocessing step, the CT scans are smoothed with an anisotropic diffusion filter. Our algorithm uses a multiscale approach: We compute an image pyramid with three different scales. The algorithm executes ten iterations on the coarsest scale, ten iterations on the second scale and finally 20 iterations on the finest scale, which corresponds to the preprocessed original image. We evaluated the influence of β by experimenting with different values. Finally, we repeated our tests with an ASM with the same configuration, but with standard shape constraints.

For assessing segmentation quality, we computed the measures Volumetric Overlap Error (VOE), Average Symmetric Surface Distance (ASD), Root Mean Square Symmetric Surface Distance (RMS) and Maximum Symmetric Surface Distance (MD). Detailed descriptions of these measures can be found in the evaluation study by Heimann et al..[15]

4. RESULTS

The quantitative results of our experiments are shown in Table 1. One can see that our new approach significantly reduces the VOE, ASD and RMS of the obtained segmentations to the ground truth in comparison to ASM segmentation with standard shape constraints. This result is also illustrated in Figure 1, which shows qualitative segmentation results on one of our test images.

On the first sight, the influence of the texture energy parameter β may appear small, because the average segmentation accuracy is roughly the same. However, when we compared individual results, we observed a strong effect of β: On some images the difference between the ASDs of the segmentations is larger than one millimeter. In general, we can see in Table 1 that the standard deviation of the results is smaller for small values of β. Moreover, we can observe that the boundary of the segmentation becomes less smooth when high values for β

Table 1. *Left*: Segmentation results of our ASM with different parameter values for β, as well as results obtained with the original ASM with default shape constraints. The numbers show average results ± standard deviation.

	VOE [%]	ASD [mm]	RMS [mm]	MD [mm]
Our ASM, $\beta = 4$	11.3 ± 4.7	2.1 ± 1.3	4.0 ± 2.3	28.3 ± 13.9
Our ASM, $\beta = 5$	11.1 ± 4.3	2.1 ± 1.2	4.1 ± 2.3	29.1 ± 13.9
Our ASM, $\beta = 6$	11.1 ± 4.8	2.1 ± 1.4	4.1 ± 2.7	29.4 ± 15.7
Original ASM	14.6 ± 4	2.7 ± 1.1	4.6 ± 2.1	29.3 ± 12.8

are used. This is illustrated in Figure 3, which compares the segmentations of one image obtained with different texture energies.

Our algorithm requires between three to five minutes to segment a liver scan. Especially preprocessing and liver detection is highly dependent on the image size. The anisotropic diffusion filter needs between 20 seconds to two minutes for smoothing the image. The initialization requires between ten seconds to one minute. The running time of the ASM algorithms themselves is almost constant. The standard ASM runs approximately 90 seconds, and our new ASM with energy minimization approximately 100 seconds.

5. DISCUSSION

We presented a new energy minimization approach for constraining shapes in ASM segmentation. Furthermore, we incorporated several novel ideas in the segmentation algorithm we used during algorithm. In particular, we proposed an approach for detecting the liver and an initialization method for the SSM with a modified ICP algorithm. Moreover, we employ a kNN-classifier based appearance model with fuzzy class labels.

We evaluated our method for constraining shapes on a set of 34 liver CT scans and compared it with standard shape constraints. In our experiments, we could observe that our method decreases the VOE, RMS and ASD significantly, while the MD remains approximately the same. We think that the reason for the better RMS and ASD errors is that our internal energy allows points to deviate from the learned PCA subspace, such that they can move closer to the surface boundary detected by the appearance model. Our current approach has a single parameter which controls the force of the external energy. Our experiments show that on some images, a higher texture energy pays off, whereas on others, it decreases the segmentation accuracy. Our interpretation of this is that a high texture energy tends to lead to higher delineation accuracy whenever the appearance model finds the true boundary. On the downside, whenever the appearance model does not detect the true boundary, a high texture energy leads to segmentation results that are not smooth, because the texture energy overrules the shape energy.

Some related SSM-based approaches for liver segmentation in contrast enhanced CT images[4,15,16] report slightly better average segmentation results. One explanation for this might be that most of these approaches[2,4,16] use a larger training data base. While adding more training data is one possible way to improve the accuracy, we still see large potential for further improvement of our method itself:

- Our evaluation shows that balancing image and model energy is crucial for our approach. In this paper, we focused mainly on the shape energy and used a relatively simple texture energy term. We are convinced that a more sophisticated texture energy term is a key for getting more accurate segmentations.
- Our shape energy has the appeal of being completely parameter free. However, our experiments showed that high texture energy can lead to segmentation results that are not smooth. Therefore, it can be beneficial to add a regularization term to the shape energy which enforces smooth segmentations. This can additionally prevent folding of triangles, which can potentially occur in the original SSM. Moreover, adding a regularization constant to the covariance matrix[6] can be used to avoid over-fitting, which is typically a problem in case of small training data sets.
- By replacing the linear shape energy term by a nonlinear energy, for example based on kernel PCA,[17] it may become possible to build SSMs for shape classes for which the assumption of Gaussian distributed shapes does not produce satisfactory results.

The major advantage of our new ASM algorithm over hybrid approaches is that we use a purely statistically motivated and theoretically sound SSM. It supports optimizing the shapes in the original data space instead of working in a linear subspace, and thus allows the segmentations to leave the SSM's PCA subspace. Many other approaches[2–4] achieve the latter only through a post-processing operation, which requires additional model assumptions and introduces additional parameters such as tension and rigidity forces.[3] Because we can still use dimensionality reduction, optimization in the original space can be done efficiently even for large 3D SSMs. Although we evaluated our algorithm on CT scans of the liver, the method is general and can be used for segmenting other organs as well. Our model is a natural extension of the original SSM,[1] which can be easily integrated in every SSM-based segmentation system.

REFERENCES

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding* **61**(1), pp. 38–59, 1995.

[2] D. Kainmueller, T. Lange, and H. Lamecker, "Shape constrained automatic segmentation of the liver based on a heuristic intensity model," in *Proc. MICCAI Workshop on 3D Segmentation in the Clinic*, pp. 109–116, 2007.

[3] T. Heimann, S. Münzing, H.-P. Meinzer, and I. Wolf, "A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation," in *Information Processing in Medical Imaging*, pp. 1–12, 2007.

[4] M. Erdt, M. Kirschner, S. Steger, and S. Wesarg, "Fast automatic liver segmentation combining learned shape priors with observed shape deviation," in *IEEE International Symposium on Computer-Based Medical Systems*, pp. 249–254, 2010.

[5] J. Koikkalainen, T. Tölli, K. Lauerma, K. Antila, E. Mattila, M. Lilja, and J. Lötjönen, "Methods of artificial enlargement of the training set for statistical shape models," *IEEE Transactions on Medical Imaging* **27**(11), pp. 1643–1654, 2008.

[6] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr, "Diffusion snakes: Introducing statistical shape knowledge into the mumford-shah functional," *International Journal of Computer Vision* **50**(3), pp. 295–313, 2002.

[7] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), pp. 239–256, 1992.

[8] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, pp. 696–710, 1997.

[9] K. Li, X. Wu, D. Chen, and M. Sonka, "Optimal surface segmentation in volumetric images-a graph-theoretic approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(1), pp. 119–134, 2006.

[10] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming B* **45**, pp. 503–528, 1989.

[11] L. Ruskó, G. Bekes, G. Németh, and M. Fidrich, "Fully automatic liver segmentation for contrast-enhanced CT images," in *Proc. MICCAI Workshop on 3D Segmentation in the Clinic*, pp. 143–150, 2007.

[12] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A: Optics, Image Science, and Vision* **4**(4), pp. 629–624, 1987.

[13] B. van Ginneken, R. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Transactions on Medical Imaging* **21**(8), pp. 924–933, 2002.

[14] M. Kirschner and S. Wesarg, "Construction of groupwise consistent shape parameterizations by propagation," in *Proc. SPIE Medical Imaging 2010: Image Processing*, 2010.

[15] T. Heimann, B. van Ginneken, M. Styner, and et al., "Comparison and evaluation of methods for liver segmentation from CT datasets," *IEEE Transactions on Medical Imaging* **28**, pp. 1251–1265, 2009.

[16] H. Ling, S. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and D. Comaniciu, "Hierarchical, learning-based automatic liver segmentation," in *Computer Vision and Pattern Recognition 2008*, pp. 1–8, 2008.

[17] D. Cremers, T. Kohlberger, and C. Schnörr, "Shape statistics in kernel space for variational image segmentation," *Pattern Recognition* **36**(9), pp. 1929–1943, 2003.