

Active Shape Models for a Fully Automated 3D Segmentation of the Liver – An Evaluation on Clinical Data

Tobias Heimann, Ivo Wolf, and Hans-Peter Meinzer

Medical and Biological Informatics, German Cancer Research Center, Heidelberg
t.heimann@dkfz.de

Abstract. This paper presents an evaluation of the performance of a three-dimensional Active Shape Model (ASM) to segment the liver in 48 clinical CT scans. The employed shape model is built from 32 samples using an optimization approach based on the minimum description length (MDL). Three different gray-value appearance models (plain intensity, gradient and normalized gradient profiles) are created to guide the search. The employed segmentation techniques are ASM search with 10 and 30 modes of variation and a deformable model coupled to a shape model with 10 modes of variation. To assess the segmentation performance, the obtained results are compared to manual segmentations with four different measures (overlap, average distance, RMS distance and ratio of deviations larger 5mm). The only appearance model delivering usable results is the normalized gradient profile. The deformable model search achieves the best results, followed by the ASM search with 30 modes. Overall, statistical shape modeling delivers very promising results for a fully automated segmentation of the liver.

1 Introduction

The computerized planning of liver surgery has an enormous impact on the selection of the therapeutic strategy [1]. Based on pre-operative analysis of image data, it provides an individual impression of tumor location, the exact structure of the vascular system and an identification of liver segments. The additional information can potentially be life-saving for the patient, since anatomical particularities are far easier to spot in a 3D visualization. The limiting factor to utilize operation planning in clinical routine is the time required for the segmentation of the liver, an essential step in the planning workflow which takes approximately one hour with conventional semi-automatic tools. For this reason, there have been numerous attempts to automate the segmentation process as much as possible.

Soler et al. have presented a framework for a complete anatomical, pathological and functional segmentation of the liver [2], i.e. including detection of the vascular system and lesions. The method is based on a shape constrained deformable model [3], which is initialized with a liver template shape and deformed by a combination of global and local forces. Park et al. use an abdominal probabilistic atlas to support voxel classification in CT images [4]. The underlying

classifier is based on a Gaussian tissue model with Markov Random Field regularization. Lamecker et al. have built a 3D statistical shape model of 43 segmented livers and utilize a modified Active Shape search for segmentation [5]. The point correspondences for their model are determined by a semi-automatic geometric method. It is interesting to note that all these approaches use some kind of shape information to guide the segmentation process: Due to its close proximity to organs with similar gray-value and texture properties, segmentation methods without prior information are prone to fail. At the same time, modeling shape information of the liver poses a huge challenge, since the enormous anatomical variance is hard to capture. Statistical Shape Models as introduced by Cootes et al. [6] seem to be best-suited for this task: In contrast to an anatomical atlas or a deformable model they do not only store information about the expected mean shape, but also about the principal modes of variation.

Since Lamecker et al. evaluated a statistical shape model for liver segmentation, there have been several advances in 3D model building and search algorithms. The time seems ripe for a new evaluation of the approach, which we are going to present in this paper.

2 Preliminaries

2.1 Statistical Shape Models

Statistical Shape Models capture shape information from a set of labeled training data. A popular method to describe these shapes are point distribution models [6], where each input shape is specified by a set of n landmarks on the surface. Applying principal component analysis to the covariance matrix of all landmarks delivers the principal modes of variation \mathbf{p}_m in the training data and the corresponding eigenvalues λ_m . Restricting the model to the first c modes, all valid shapes can be approximated by the mean shape $\bar{\mathbf{x}}$ and a linear combination of displacement vectors. In general, c is chosen so that the model explains a certain amount of the total variance, usually between 90% and 99%. In order to describe the modeled shape and its variations correctly, landmarks on all training samples have to be located at corresponding positions.

2.2 Gray-Level Appearance Models

To fit the generated shape model to new image data, a model of the local appearance around each landmark in the training data is necessary. For this purpose, Cootes et al. suggest to sample profiles perpendicular to the model surface at each landmark [7]. Typically, these profiles contain the immediate gray-level values or their normalized derivatives. By collecting profiles from all training images, a statistical appearance model featuring mean values and principal variations can be constructed for each landmark. The probability that an arbitrary sample is part of the modeled distribution can then be estimated by the Mahalanobis distance between the sample and the mean profile.

2.3 Active Shape Model Search

Starting from the mean shape and an initial global transform to the image, the shape model is refined in an iterative manner: First, the fit of the gray-level appearance model is evaluated at several positions along the surface normal for each landmark. Subsequently, the global transform and the shape parameters y_m are updated to best match the positions with the lowest Mahalanobis distance. To keep the variation in reasonable limits, the shape parameters are restricted either to $\pm 3\sqrt{\lambda_m}$ individually or to a hyperellipsoid for the entire shape vector \mathbf{y} . By repeating these steps, the model converges toward the best local fit. To make the method more robust regarding the initial position, usually a multi-resolution framework is employed [7]: The model is first matched to coarser versions of the image with a larger search radius for the new landmark positions.

3 Material and Methods

3.1 Image Data

The data used in the experiments has been collected over a period of five years of computerized operation planning and clinical studies at our research center. All images are abdominal CT scans enhanced with contrast agent and recorded in the venous phase, though the exact protocol used for acquisition changed over time. The resolution of all volumes is 512x512 voxels in-plane with the number of slices varying between 60 and 130. The voxel spacing varies between 0.55mm and 0.8mm in-plane, the inter-slice distance is mostly 3mm with a few exceptions recorded with 5mm slice distance.

From 99 scans that have been labeled by radiologic experts, eight had to be taken out of the experiments because of abnormal anatomy, e.g. in cases where parts of the liver have been removed by surgery or where tumors exceeding the volume of one liter deform the surroundings. The quality of the individual segmentations, created with a selection of different manual and semi-automatic tools, varies depending on the application they were created for. For some datasets the V.Cava is segmented as part of the liver, for others it is left out. We wanted to build a model without the V.Cava (as it is used for operation planning) and elected 32 datasets with high quality segmentations as training samples for the shape model. The chosen segmentations were smoothed with a 3D Gaussian kernel and converted to a polygonal representation by the Marching-Cubes algorithm. The resulting meshes were then decimated to contain around 1500–2000 vertices and forwarded to the model-building process. The remaining 59 CT volumes were treated as candidates for the evaluation process.

3.2 Model Building

While Lamecker et al. use a semi-automatic procedure to determine correspondences, we employ a fully automated approach that minimizes a cost function

based on the description length of the model [8], which should deliver better generalization ability and specificity. The shape model was built with 2562 landmarks which were distributed equally over the model surface employing the landmark reconfiguration scheme we recently presented in [9]. For the gray-level appearance models, a multi-resolution image pyramid was created for each CT volume that was used during the creation of the shape model. We opted for six different levels of resolution R_0 to R_5 , where R_n corresponds to the n -th down-sampling step (with R_0 as the original resolution). For each down-sampling step, x - and y -resolution were halved, leading to a doubled voxel spacing. When the xy -spacing reached the same values as the (originally much higher) z -spacing, the z -resolution was halved as well. Finally, we calculated three different gray-level appearance models for each resolution: A plain intensity profile, a gradient profile and a normalized gradient profile.

3.3 Evaluation of Gray-Level Appearance Models

To evaluate the performance of the different appearance models, we employ the following procedure: For all training images, the fit of the gray-level appearance models is evaluated at the true landmark positions and at three positions on each side of the surface. To simulate the conditions during model search (where landmarks are most probably not located at the correct positions), we randomize the landmark position with a standard deviation of 1mm in R_0 along the surface (doubled at each following resolution). At the same time, the direction of the normal vector is randomized with a standard deviation of approximately ten degrees. This way, 20 samples are extracted for each landmark in each image. The index of the position with the best fit (ranging from -3 to 3) is stored and used to generate a histogram of the displacements for each resolution. Ideally, the appearance model should always achieve the best fit at the true position (displacement 0), in practice we expect to see a Gaussian distribution with a certain variance.

3.4 Alternative Model Search Algorithm

In the classical ASM approach, the model is strictly constrained to the variations learned from the training data. To allow additional degrees of freedom, Weese et al. presented a search method with shape constrained deformable models [10]. They calculate the external energy from the fit of gray-value profiles (similar to the original ASM search) and the internal energy based on the length difference between corresponding edges in the deformable mesh and the closest shape model. A conjugate gradient algorithm is used to minimize the weighted sum of both energies, varying the position of all vertices.

We adopt the idea of allowing free deformations guided by the difference in edge length, but simplify the method for an easier integration into the ASM search algorithm as described in Sec. 2.3: When the new landmark candidates are found, a spring-model deformable mesh is initialized on these points with the neutral positions for all springs set to the corresponding edge length of the

closest shape model. This mesh is then iteratively relaxed according to:

$$x_i = x_i + \sum_{j \in N(i)} d_{i,j} \delta \frac{|d_{i,j}| - |m_{i,j}|}{|d_{i,j}|} \quad (1)$$

where x_i are the coordinates of the i th point, $N(i)$ denotes the neighbours of vertex i , and $d_{i,j}$ and $m_{i,j}$ are directed edges of the deformable mesh and the model, respectively. δ is set to 0.05 in our experiments, and a number of 100 iterations is run for relaxation.

3.5 Evaluation of Model Search

Initially, the shape model is scaled to a fixed size (around the average liver size) and translated to the upper left part of the image volume (corresponding to the right side of the patient). Without any further interaction, this procedure leads to the model being attracted by the liver in the vast majority of cases. In 11 cases, however, the image volume was recorded at a different patient position or slightly rotated, so that the search would not converge towards the liver. Instead of devising special initial transforms for these images, we decided to drop them from the evaluation set and ended up with 48 volumes for the final validation.

Initial experiments suggested that the best starting resolution for the search is R_4 , since many profiles leave the image volume in R_5 , reducing the information of the appearance model. We run a fixed number of 10 iterations for R_4 and R_3 each, which usually brings the model pretty close to the liver. To deal with the remaining cases, the search in R_2 is run until convergence, which is defined as a maximum landmark movement of 0.5mm. Subsequently, 10 iterations in R_1 and R_0 each are sufficient to fine-tune the model to the image data. This method was performed with three different search strategies: The ASM search with 10 modes of variation for the model (1), the ASM search with an increased 30 modes of variation in R_2 to R_0 (2) and the deformable model search with 10 modes of variation (3). In the latter case, the deformable model was only used in R_1 and R_0 , the previous resolutions were handled as in method 1. For all methods, the shape parameter vector \mathbf{y} was restricted to lie inside a hyperellipsoid (size determined by the χ^2 distribution). Originally, we planned to evaluate a combination of these methods with all created appearance models, but it quickly became evident that only the normalized gradient appearance model delivered usable results (see Sec. 4.2).

After the last iteration in R_0 terminates, the resulting mesh is rasterized into a volume of the same resolution and compared to the existing segmentation. A number of different comparison measures is used for this purpose: The Tanimoto coefficient which quantifies the overlap of two volumes as $C_T = |A \cap B| / |A \cup B|$, average and RMS surface distance and the ratio of the surface area with a deviation larger than 5mm. All surface metrics were calculated in both direction to guarantee symmetry.

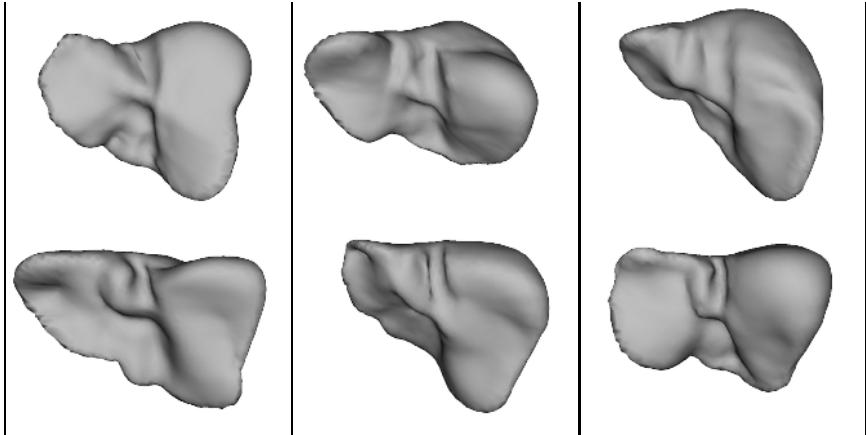


Fig. 1. Visualization of the variance of the created shape model: The left column shows the variation of the largest eigenmode between $\pm 3\sqrt{\lambda_1}$, the medium and right column of the second and third largest eigenmode, respectively

4 Results

4.1 Model Building

For a detailed evaluation of the model building process, we refer the reader to [9]. The three largest modes of variation are displayed in Fig. 1 and seem to capture the encountered shape variability adequately. 90 percent of the total variance is explained by the first 10 modes of variation (used for search methods 1 and 3), while the 30 modes used for method 2 account for 99.9 percent.

4.2 Evaluation of Gray-Level Appearance Models

The results of the displacement from true position analysis are displayed in Fig. 2. While we expected a Gaussian distribution for all appearance models, only the normalized gradient profile produces symmetric displacements. In contrast, the plain intensity and gradient profiles feature a clear shift to the inside of the shape model (negative displacement values).

4.3 Evaluation of Model Search

Figure 3 displays boxplots of the results of the automatic segmentation using the three different search techniques. The results of the volumetric error are specified as $100(1 - C_T)$ (C_T being the Tanomoto coefficient). For all four measures of segmentation quality, the ASM search with 30 modes of variance yields better results than the search with 10 modes. For one test dataset, the search with 30 modes did not converge in R_2 , this image was omitted from the statistics of method 2. The best overall results are accomplished with the deformable model search.

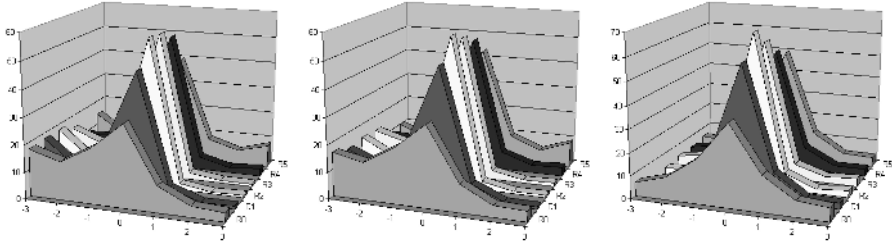


Fig. 2. Histograms showing the displacements from the true landmark positions at different resolutions R_0 to R_5 . From left to right: Intensity, gradient and normalized gradient profile appearance models.

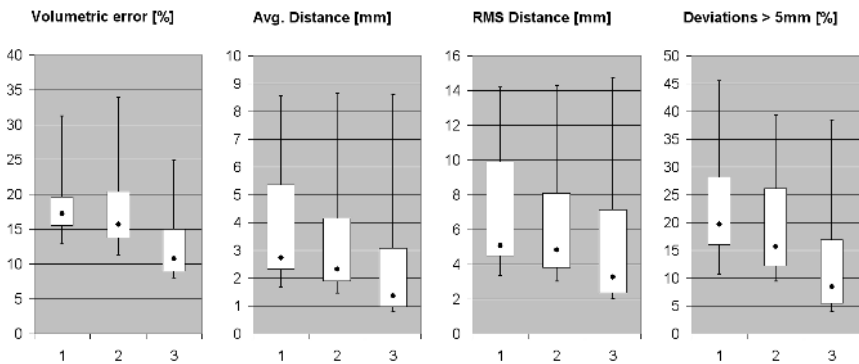


Fig. 3. Results of the segmentation using the normalized gradient appearance model (1=ASM with 10 modes, 2=ASM with 30 modes, 3=deformable model with 10 modes). The box connects the 1st and 3rd quartiles of all values with the dot representing the median, the whiskers span the interval between the 0.05 and 0.95 quantiles.

5 Discussion

Comparing our results to the ones obtained in [5] (2.3–3.0mm average surface distance, 3.1–3.9mm RMS distance and 9.0–17.1% deviations larger 5mm for a varying number of parameters during model search) does not reveal significant differences. However, it is hard to draw any conclusions from this, mainly because different training and evaluation data was used in the experiments. Consequently, the here presented numbers should only be interpreted relative to each other.

Having evaluated our shape model on nearly 50 clinical datasets, we are confident that a statistical shape modeling approach is able to solve the segmentation problem for liver operation planning for the vast majority of cases in the near future. However, we also noticed several problems: While 32 training shapes do not build the most extensive shape model, it is a sufficiently high number to draw the conclusion that the necessary shape variability for an exact segmentation of the liver will probably not be reached by the strictly constrained ASM

model. Approaches using deformable meshes based on the shape model seem to have a much higher potential of solving this task. Considering the simplicity of the deformable model enhancement, the obtained improvements for already acceptable results of the ASM are excellent. Since the deformable model was only used in the two highest resolutions, it could not save the performance in the worst case results, as is noticeable by the upper whiskers in the boxplots. A better initialization of the shape model and restriction of the allowed geometric transformations (rotation and scale) seem to be necessary in these cases. We were surprised by the disappointing results from the intensity and unnormalized gradient appearance models, which are probably due to the skew distribution visible in the displacement histograms. However, there are many more possibilities to model the local appearance (e.g. [11]) which will most likely improve the obtained results. Our future work will focus on evaluating these alternative appearance models and on improving the deformable model search with more sophisticated search and relaxation schemes.

References

1. Meinzer, H.P., Thorn, M., Cardenas, C.E.: Computerized planning of liver surgery – an overview. *Computers & Graphics* **26** (2002) 569–576
2. Soler, L., Delingette, H., Malandain, G., Montagnat, J., et al.: Fully automatic anatomical, pathological, and functional segmentation from ct scans for hepatic surgery. In: *Proc. SPIE Medical Imaging*. (2000) 246–255
3. Montagnat, J., Delingette, H.: Volumetric medical images segmentation using shape constrained deformable models. In: *CVRMed*. (1997) 13–22
4. Park, H., Bland, P.H., Meyer, C.R.: Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE TMI* **22** (2003) 483–492
5. Lamecker, H., Lange, T., Seebass, M.: Segmentation of the liver using a 3D statistical shape model. Technical report, Zuse Institute, Berlin (2004)
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models – their training and application. *CVIU* **61** (1995) 38–59
7. Cootes, T.F., Taylor, C.J.: Statistical models of appearance for computer vision. Technical report, Wolfson Image Analysis Unit, University of Manchester (2001)
8. Heimann, T., Wolf, I., Williams, T.G., Meinzer, H.P.: 3d active shape models using gradient descent optimization of description length. In: *Proc. IPMI, Springer* (2005) 566–577
9. Heimann, T., Wolf, I., Meinzer, H.P.: Optimal landmark distributions for statistical shape model construction. In: *Proc. SPIE Medical Imaging*. (2006) 518–528
10. Weese, J., Kaus, M., Lorenz, C., Lobregt, S., et al.: Shape constrained deformable models for 3D medical image segmentation. In: *Proc. IPMI, Springer* (2001) 380–387
11. van Ginneken, B., Frangi, A.F., Staal, J.J., ter Haar Romeny, B.M., Viergever, M.A.: Active shape model segmentation with optimal features. *IEEE TMI* **21** (2002) 924–933