Visual Computer manuscript No.
(will be inserted by the editor)

Opening up the "black box" of Medical Image Segmentation with Statistical Shape Models

Tatiana von Landesberger · Gennady Andrienko · Natalia Andrienko · Sebastian Bremm · Matthias Kirschner · Stefan Wesarg · Arjan Kuijper

Abstract The importance of medical image segmentation increases in fields like treatment planning or computer aided diagnosis. For high quality automatic segmentations, algorithms based on statistical shape models (SSMs) are often used. They segment the image in an iterative way. However, segmentation experts and other users can only asses final results, as the segmentation is performed in a "black box manner". Users cannot get deeper knowledge on how the (possibly bad) output was produced. Moreover, they do not see whether the final output is the result of a stabilized process.

We present a novel Visual Analytics method, which offers this desired deeper insight into the image segmentation. Our approach combines interactive visualization and automatic data analysis. It allows the expert to assess the quality development (convergence) of the model both on global (full organ) and local (organ areas, landmarks) level. Thereby, local patterns in time and space, e.g., non-converging parts of the organ during the segmentation, can be identified. The localization and specifications of such problems helps the experts creating segmentation algorithms to identify algorithm drawbacks and thus to point them to possible ways how to improve the algorithms systematically.

We apply our approach on real-world data showing its usefulness for the analysis of the segmentation process with statistical shape models.

T. von Landesberger, S. Bremm, M. Kirschner
TU Darmstadt, Darmstadt, Germany
E-mail: name.surname@gris.tu-darmstadt.de

N. Andrienko, G. Andrienko
Fraunhofer IAIS, Bonn, Germany
E-mail: name.surname@iais.fraunhofer.de

A. Kuijper, S. Wesarg
Fraunhofer IGD, Darmstadt, Germany
E-mail: name.surname@igd.fraunhofer.de

Keywords Medical Imaging · Medical Modeling · Visual Analytics · Image segmentation · Statistical Shape Models · Spatio-Temporal Data

1 Introduction

Segmentation of 3D medical images serves as an indispensable basis in many medical disciplines and diagnostic processes. Manual segmentation is very costly and time consuming as it needs to be performed by medical experts in a piecewise manner: each 2D image slice is examined individually. Thus, there is a strong need for automatic segmentation algorithms [19].

Commonly used automatic segmentation methods are based on statistical shape models (SSMs). They are robust and can cope with low contrast images [12]. These algorithms work iteratively. After the initial positioning of the organ model, its landmarks are progressively moved so that they fit onto the real data. Usually, the calculation finishes after a pre-defined number of iterations.

Although there are many well-accepted SSM-based algorithms [12], they can still perform poorly on some data samples. It is then relevant to understand *why* and *how* the output was produced. Unfortunately, the modeling experts (also referred to as users) can assess only the final output without the possibility to analyze the *progress* of the segmentation (i.e., the "black box").

We present a novel Visual Analytics method for assessing SSM-based segmentation. This new method allows to analyze *global* and *local* segmentation performance. We thereby differentiate between two types of performance (i.e., convergence): quality development and stabilization of output. This is deemed relevant as there might be cases when the general quality improves

but the output varies significantly with a small change in the number of iterations (see Fig. 1).

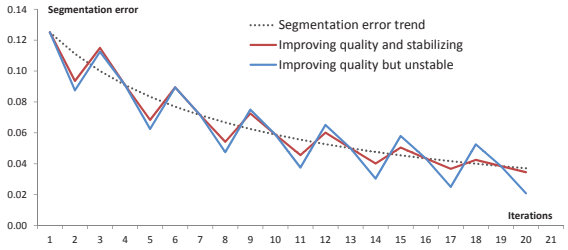

Fig. 1 Illustration of different convergence patterns according to the stabilization criteria.

With help of our method, the user can answer the following questions:

- Global convergence:
  - Quality development: Does the global segmentation quality improve over the segmentation period? Should the number of iterations be changed?
  - Stabilization: Does the segmentation converge to a stable output (i.e. are the changes in quality decreasing)?
- Local convergence:
  - Quality development: What are the typical and atypical convergence paths (movements) of landmarks during segmentation?
  - Quality development: Are there areas or organ landmarks that deteriorate in quality during segmentation phase?
  - Stabilization: Are there areas that are unstable (i.e., do landmarks significantly move also at the end of the process)?

Our method is based on a combination of interactive visualization and automated data analysis. Several linked views show the segmentation progress, provide an overview of the types of movements of organ landmarks during the segmentation, and highlight the organ areas with characteristic convergence patterns. We employ self organizing maps (SOM) for the clustering and visualization of landmark movements. For this purpose, we introduce new features describing the landmark movements with regard to the segmentation convergence (quality development and stabilization).

The new method enables experts creating segmentation algorithms to gain deeper understanding of the segmentation process. It helps them to detect problems of current algorithms and thereby to create better models. This will in turn also improve medical treatment planning and diagnosis.

The paper is structured as follows: Section 2 introduces SSM-based image segmentation. Section 3 reviews related work. Section 4 presents our approach for the Visual Analytics support of the segmentation process assessment. Section 5 demonstrates the application of our approach on real-wold data. Section 6 discusses our approach and, finally, Section 7 concludes and outlines future work.

2 Image Segmentation Based on Statistical Shape Models

Medical image segmentation based on statistical shape models consists of three main stages (see Fig. 2):

Fig. 2 SSM-based image segmentation. After the model specification, model-based segmentation is performed. Finally, the segmentation result is evaluated.

1. *Model specification*: Two parts of the model are specified: the shape model and the appearance model.
2. *Image segmentation:* The selected model and input data are used for iterative segmentation.
3. *Segmentation evaluation*: The segmentation is evaluated using ground truth data.

We next discuss statistical shape models and describe the image segmentation phase in more detail.

2.1 Statistical Shape Models

Statistical Shape Models (SSMs) represent the typical *shape* of an object class and likely shape variations by means of a probability distribution. The distribution is learned from a set of training shapes [12]. Each shape is represented with a fixed number of corresponding points, called *landmarks*. The variability of the data (i.e., the data model), is in most cases captured by a principal component analysis (PCA) of these high-dimensional training shapes.

In addition, SSM-based segmentation algorithms like the Active Shape Model [8] employ a *local appearance model*, which describes the texture around the organ boundary in input images. The appearance models are included either implicitly (e.g., by searching for strongest edges), or are learned from training data.

2.2 Image Segmentation Phase

SSM segmentation is usually done with the Active Shape Model [8] (see Fig. 3). The algorithm is initialized by

placing the model's mean shape roughly onto the target organ in the scan, which can be done either interactively or by using an automatic object detection algorithm. After initialization, the shape is iteratively adapted to the image. Each iteration consists of two main steps: The first step ("AM") comprises two parts. First, the organ boundary is searched using a local search around each landmark. For each landmark, a set of candidate image features is sampled and the quality of each candidate is assessed by the landmark's local appearance model. Then, one image feature is chosen for each landmark, and the landmark is displaced to the feature's position. In our implementation, this feature selection is done by solving a max-flow-problem as proposed by [13], which ensures that the displacement of neighboring landmarks is consistent. In the second step ("SM"), the shape constraints are applied, so that the shape is consistent to the learned model. The algorithm finishes after a fixed number of iterations.

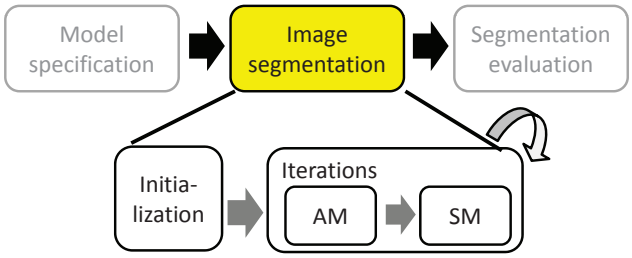

Fig. 3 Image segmentation phase consists of an initialization and a predefined number of iterations, consisting of two steps based on appearance "AM" and shape constraints "SM".

Figure 4 shows example results of the two steps based on appearance "AM" and shape constraints "SM". SM step creates a smooth organ boundary conform with the learned shape variability. The SM step is the final outcome of the iteration. In the following, if not stated otherwise, we use it in the examples.

3 Related Work

Visual Analytics techniques for medical imaging currently form only a small part of the wide portfolio of Visual Analytics methods. They apply automatic data analysis methods for highlighting interesting views on the data or for guiding the interaction with the visualization, or show additional data in linked views (e.g., [4,5,19]). They do not only combine data analysis and interactive visualization, but also bridge information visualization and scientific visualization. However, this research field is still in a premature stage. Currently, there are few visual analytics methods for support of SSM-based medical image segmentation.

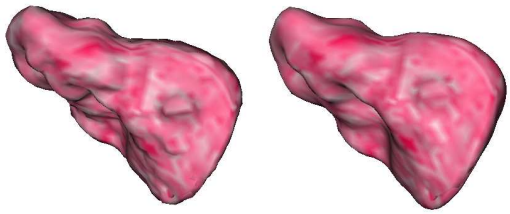

Fig. 4 Current segmented organ in various substeps. Left: Appearance-based step, right: Shape-based step.

Busking et al. [7] developed a technique for visualization of the *shape space*. They present several visualizations for this purpose (PCA reduction, shape overlap, etc.). This allows for assessing the distribution of the input data from various perspectives. This approach however only shows the distribution of the input samples and does not support analysis of the segmentation phase and the evaluation of the segmentation results.

There are two main ways of *segmentation evaluation*: algorithmic (i.e., quantitative) and visual. Currently, the evaluation techniques focus only on the assessment of the final output. In the quantitative evaluation, the results are compared to manual segmentation using a few global metrics [11]. Visual inspection commonly employs standard visualization techniques: 2D and 3D views. ITK Snap (`www.itksnap.org`) shows user-selected 2D slices of the real data overlaid with the resulting segmentation. In 3D, methods for visualization of differences between two surface meshes (corresponding to the expert and automatic segmentation) overlay the two meshes using transparency and additional visual cues [6,10,22,26]. Such visual inspection is subjective and time-consuming owing to a large amount of available views that have to be examined. Moreover, they show solely the result without the segmentation process.

Usually, the *image segmentation phase* is performed in a "black-box" manner, which does not allow the user to track the transformation from input to output. Depending on the choices in the initial analysis step, the results may emerge that do not comply with user preferences or the application context. In these cases, it is advantageous to allow the user to interactively oversee and control the model adaptation process. Such a technique has been introduced for clustering of 2D trajectories of financial data [21]. It, however, cannot be applied to the analysis of the segmentation phase in medical context.

The positions of organ landmarks during the iterations of the segmentation phase form paths – trajectories (see Sec. 4). So the assessment of the segmentation on a local level is somewhat connected to analysis of spatio-temporal data, in particular movements. In geographic visualization, there are numerous methods for visual analysis of movements [3]. Relevant to our work are methods for clustering trajectories by similarity and visualizing cluster profiles [20,21], methods for continuous [25] and discrete aggregation [2] of trajectories, and methods that cluster time intervals by similarity of spatial situations representing movement [1].

In most cases, these methods are however restricted to movements in 2D space and, most importantly, they do not focus on analysis of convergence - whether the organ landmarks head towards the ground truth.

In spatio-temporal modeling, several papers report about interactive visualization designed to help users to *explore and evaluate models*. Thus, Demšar et al. [9] employ coordinated linked views and clustering for exploration of a geographically weighted regression model of spatio-temporal data. Matković et al. [17] support users in exploring multiple runs of a simulation model. Maciejewski et al. [16] suggest visual analytics techniques to support the exploration and use of existing spatio-temporal models. The user can explore the model results on a map and time series display. These methods do not allow the user the deep insight how the models work, they only show outputs of various model runs.

In *summary*, there are few visual analytics methods for the support of SSM-based medical image segmentation and various methods for spatio-temporal analysis. However, there are no specific methods supporting analysis of the image segmentation process.

4 Approach

We present new visual methods for assessing the segmentation progress on global (full organ) and local level (landmarks or areas). Figure 5 shows an overview of our approach. It consists of three linked views supported by algorithmic data analysis.

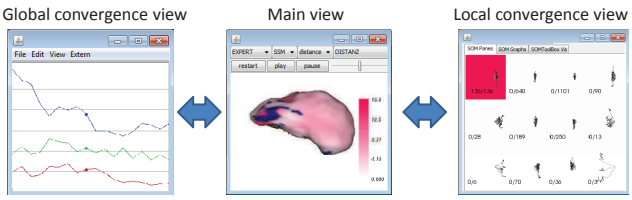

Fig. 5 Overview of our approach. The main view on segmentation progress is linked with global and local convergence criteria exploration.

1. *Global quality view*: The view shows the progress of segmentation quality based on several common quality metrics [11]. It gives a coarse view on the process convergence.
2. *Main view*: It shows both an animated view of the segmentation process and a static view on current state in a selected iteration. This central view is linked to both the global quality and the local pattern view.
3. *Local pattern view:* It shows an overview of the different types of local (per landmark) convergence patterns using self-organizing maps (SOM).

In addition, we provide a view on the correspondence between the expert segmentation and the segmentation iterations (see Section 4.1.3).

4.1 Convergence criteria

4.1.1 Definitions

As an input for our calculations, we assume a set of segmentation output meshes (i.e., outputs or segmentations) $M^{i,s}$ for each iteration $i \in I$, $I = \{0, \ldots, L\}$ and each sub-step $s \in S = \{\text{AM, SM}\}$ of the segmentation and an expert segmentation M^E. The final segmentation is the last iteration $M = M^{L,SM}$.

Each output $M^{i,\cdot}$ has N points $p_k^{i,\cdot} \in M^{i,\cdot}$ and the expert mesh M^E has N^E points. N^E does not need to match N. The movement of points during iterations $p_k^{i,\cdot}$ forms trajectories $P_k = \{p_k^{0,\cdot}, \ldots, p_k^{L,\cdot}\}$.

Distance between two points $p_1 \in M^1$ and $p_2 \in M^2$ is denoted as $d_{i,j}^{1,2}$. The distance of a point $p_i \in M^1$ to the mesh M^2 is approximated by $d_i^{1,2} = min_j(d_{i,j}^{1,2})$, which is also used in calculating common medical quality measures such as Average Surface Distance [11]. The point $p^* \in M^2$ minimizing the distance is the *target* t.

For each landmark (i.e., point) k of the segmentation outputs $M^{i,s}$, there is a corresponding target in the expert segmentation M^E denoted $t_k^{i,s}$. As there may be several such targets $\{t.\}$ for p, we select a target t so that it is stable over the iterations (i.e., $t^i = t^{(i-1)}, t \in \{t.\}$). The targets form an ordered set T_k^s. Mean point of this set is called *center target* $\overline{t_k^s}$.

4.1.2 Global Criteria

Our quality metrics are based on the commonly used global measures Q for assessing final output quality in medical image segmentation [11]: Hausdorff Distance, Average Distance, Average Square Root Distance, and Dice's Coefficient.

We are interested in the convergence of the segmentation process. For good segmentations, we would expect not only that the quality w.r.t. expert segmentation improves during segmentation but also that the quality stabilizes (i.e., differences in quality decrease) as illustrated in Fig. 1. Therefore, we propose to measure two *global convergence aspects:*

- *Development:* quality measure of each iteration w.r.t. to the expert segmentation $Q(M^{i,s}, M^E)$.
- *Stabilization:* the quality difference to the previous iteration $Q(M^{i,s}, M^{i-1,s})$, $\forall i > 0$

The global criteria show how the segmentation performs generally, but cannot show local differences, e.g., one part of the organ is well segmented but the other is quite bad.

4.1.3 Local Criteria

The *local convergence criteria* measure whether and how each landmark $p_k^{\cdot,s}$ moves towards its target $t_k^{i,s}$. We base our calculation on the notion of "target" (see Sec. 4.1.1), as it is used in the main global quality criteria (e.g., average or maximum of landmark distances to their targets).

Target point stability

During iterations, the target point may change $t_k^{i,s} \neq t_k^{i-1,s}$. We analyze these changes as the target stability is important for our local convergence criteria described below. Moreover, the target changes can serve as a rough indication of convergence. We expect for a good convergence that the points move directly towards their targets (therefore do not change the targets a lot). Additionally, if the segmentation converges to a stable result, these changes should decrease towards the end.

We therefore propose to analyze and show the user the development of target points (see Fig. 6).

The new target-based convergence criteria are:

- *Development* measures the magnitude of the target changes. We use the following criteria in combination as they complement each other.
  - Standard deviation of targets $std(t_k^{\cdot,s})$: Large values indicate large movements. Note, that the absolute value can be influenced by expert mesh cell size.
  - Number of targets $|T_k^s|$: Larger number of targets indicates larger "sidewise" movements of landmarks irrespective of mesh resolution.
  - Number of transitions $sum(id(t_k^{i,s} \neq t_k^{i-1,s}))$: Measures stability of targets in combination with number of targets. High number of transitions for few targets indicates instability.
- *Stabilization* measures whether the target changes less with increasing iterations.
  - Target stabilization index TSI: Measures whether the target movements decrease during iterations. We propose to use a weighted average distance to the target, where weights increase with iterations. We normalize the index by average distance to the target. Index values $TSI < 1$ mean stabilization.
    $TSI = \left\{ \frac{\sum_{i=0}^{L}(a^{L-i} \cdot d(t_k^{i,s}, \overline{t_k^s}))}{\sum_{i=0}^{L}(a^i)} \right\} / \left\{ \frac{\sum_{i=0}^{L}(d(t_k^{i,s}, \overline{t_k^s}))}{L} \right\}$,
    where $a < 1.0$. The value of a influences absolute value of the index, but not its comparability across points. We employ commonly used value of a in weighted average calculations $a = 0.8$.

For target stability analysis, we provide users with a view, which shows the landmarks color-coded according to the above criteria (see Fig. 6c-f). We also show the correspondence of the targets to the first iteration and target trajectories (see Fig. 6a-b). Trajectories are color-coded for better discrimination of neighboring points.

For ensuring higher target stability, and thereby also higher reliability of the center target, the user can optionally reduce the target noise. If there are points with similar distance to the target (e.g., up to 5% difference), they are used for selection of target point so that they minimize differences to the target history ($t^i = argmin(d(t, t^{i-1}))$). Although this approximation may slightly change the exact landmark distance, it ensures higher target stability needed for the following quality criteria. In general, the assessment of segmentation process is not significantly influenced.

Landmark Movement-based convergence

For deeper analysis of local convergence, we propose to analyze movements of individual landmarks P_k during segmentation. It shows, e.g., whether landmarks move towards the target, and whether their movement stabilizes. Landmarks with similar movements can form areas of good or bad convergence, which are interesting for the user.

Initial data normalization: In order to ensure comparability of the criteria across the whole organ, we propose the following data transformation (see Fig. 7).

1. *Translation*: All trajectories start in one point $p_k^{0,\cdot} = \{0, 0, 0\}$, correcting for landmark positions across the 3D organ.
2. *Scaling*: All trajectories are scaled according to the initial distance to the target center $d(p_k^{0,\cdot}, \overline{t_k^s})$. This allows to analyze convergence (i.e., progress towards target) – length of landmark displacement in relation to the target distance.

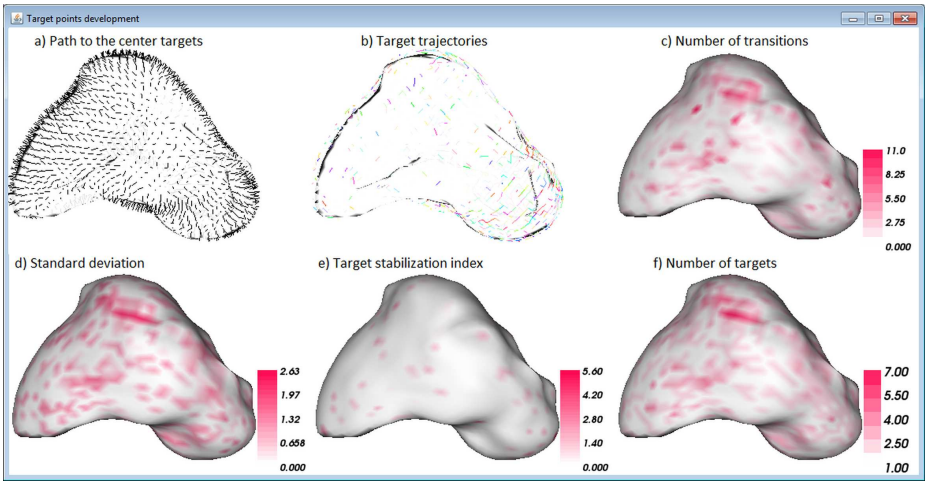

Fig. 6 View on various criteria for convergence of target points in the organ. a) Landmark targets connected to landmark positions. b) Target movements. Each target is color coded. c)-f) Various target stability criteria for landmarks shown on the organ. Values are color-coded (white to red).

3. *Rotation* All trajectories are rotated towards one common target $\overline{t} = \{1, 0, 0\}$. This corrects for various directions of targets across a 3D organ. After this step, all targets overlap in one point (see Fig. 7d).

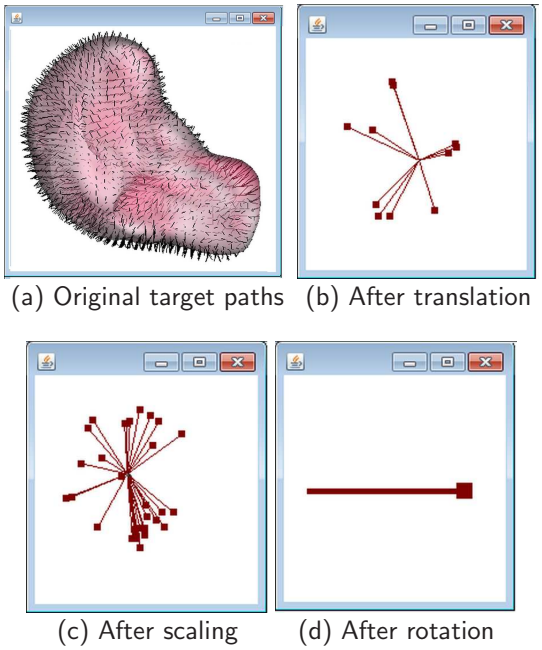

Fig. 7 Illustration of data normalization steps. After all steps, all targets are in one point (i.e., all targets overlap).

We propose new *local convergence criteria*:

- *Development*
  - *Distance to the target (D)* $d(p_k^{i,\cdot}, t_k^{i,s})$: For a good convergence, we expect the distance to decrease from 1 to 0. The distance shows how far the point is to its target, but does not capture whether the landmark moves beyond target or makes swinging movements around target (see Fig. 8).
  - *Projected Progress to the target (PP)* measures how the point behaves in relation to its optimal path. We calculate it as the projected distance to the target on the optimal path. We expect the values to increase from 0 to 1. $PP < 0$ denotes movement away from target, $PP > 1$ means movement beyond target.
- *Stabilization*
  - *Step Length* (SL) $d(p_k^{i,\cdot}, p_k^{i-1,\cdot})$: For a stable result, we expect that the size of landmark movements decreases towards 0.

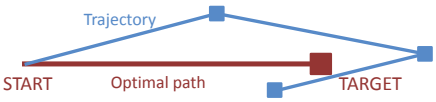

Fig. 8 Point trajectory and the optimal path.

These criteria are calculated for each iteration and therefore form a time series, or a feature vector. The values of the feature vector allow to assess the progress of the segmentation quality. For example, landmarks can only slowly move towards the target at the beginning and later head to the target faster.

4.2 Global Convergence View

The global quality view provides interactive visualization of the user-selected global criteria (see Section 4.1.2).

It can be used for examining the segmentation's convergence as well as for assessing which number of iterations is suitable for a particular data set.

Figure 9 top shows the development of Average and Average Square Root distance over all iterations and sub-iterations of a segmentation. We can clearly see a periodical pattern in the data caused by the two sub-iterations. For better assessment of the segmentation process, the user can choose to compare the sub-iterations (see Fig. 9 center).

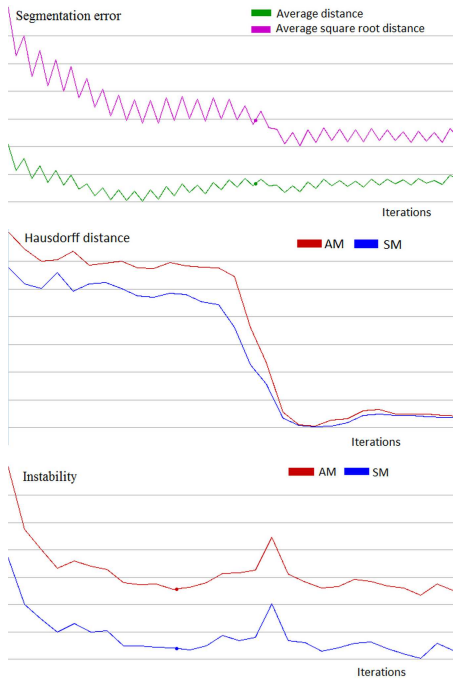

Fig. 9 Global convergence view. Top: Average and Average Square Root Distance for all steps (ALL). Center: Comparison of Hausdorff distance for AM and SM steps. Bottom: Stabilization measured by the Average Distance.

On demand, the user can also analyze the global stabilization metrics (see Fig. 9 bottom). We see that there is little stabilization, whereby the shape-based steps (SM) stabilize more than the appearance-based steps (AM).

The quality view can be also used for spotting interesting iterations (e.g., peaks, or breaks), which should be analyzed in more detail. This is supported by linking the global view with the main view.

4.3 Main View

The main view shows the segmentation iterations both using animation and trajectories.

In the animation mode, the user can play/stop the course of segmentation and thereby spot the main segmentation developments (e.g., main drifts in the organ shaping). Moreover, the user can assess the progress of local quality (distance to target). The local quality is color-coded, where red color means bad quality (see Fig. 10). The user can employ current or overall normalization. The current normalization highlights bad regions in the selected iteration and overall normalization shows quality differences over the whole segmentation period.

Animation is suitable for analyzing global trends in the data. However, owing to limited human cognitive capacities, it is difficult to analyze the developments in detail [23]. Therefore, we also provide a static view on the segmentation progress using trajectories of individual landmarks.

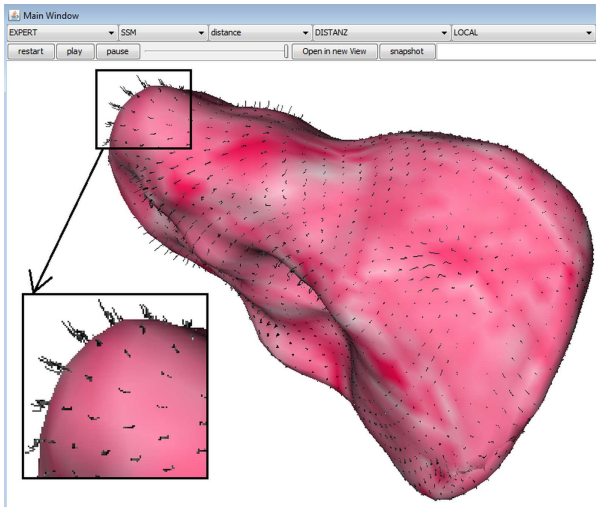

Fig. 10 Main view. Visualization of the current state of the segmentation phase. Segmentation process is shown either as animation (using controls above) or using trajectories of organ landmarks (zoomed).

The main view shows the general development of segmentation progress. However, the assessment of the data at landmark level is very cumbersome, as there is a large number of landmarks to compare (thousands). Moreover, it is not possible to compare the movements over the whole organ. Such analysis is provided in the local convergence view.

4.4 Local Convergence View

Landmark movements w.r.t. their targets are analyzed and shown in the local convergence view. The movement data consists of several thousand landmarks with dozens of iterations each. We propose to use self-organizing map (SOM) [15,24] to gain an overview of the convergence patterns in the data. We use advantageous

SOM properties: the combination of dimension reduction and clustering. SOM can handle large datasets and is amenable to visualization. Moreover, it preserves topological properties of the data (i.e., two data points that are close in original space are closely located in the SOM).

4.4.1 SOM Calculation

The SOM algorithm takes multidimensional data called "feature vectors" as an input. In our case, these feature vectors are formed by the local convergence criteria for each landmark (see Section 4.1.3). The feature vectors describe the similarity of landmarks. The result of the SOM calculation is a 2D grid of cells, each containing objects (landmarks) with high similarity to each other.

Our implementation employs the SOM Toolbox [18]. The main parameterization required by the algorithm includes the initialization of the prototype vectors and the specification of the learning parameters. We use "rules of thumb" for the parameter setting [15,24]. The calculation in our use case lasts ca. 30 seconds.

4.4.2 Views

We present the user the resulting SOM grid in two types of views: one showing the landmark trajectories and the second displaying the convergence values (see Fig. 11 and 12). The two views are used as complements.

The *SOM trajectory view* shows the trajectories of landmarks in the cells. In order to avoid strong overplotting, we show only the most representative trajectories for each cell (i.e., trajectories closest to the cell center). The number of representatives is user defined. The number of objects in each cell is shown in lower left corner of each cell. On demand, the user can see the trajectories of a selected cell in a larger view.

The *SOM feature view* offers the user a deeper understanding of the SOM result by showing the underlying features as line charts. The user can assess the exact convergence pattern. The central pattern of the cell is shown in red and patterns of the cell objects are shown in blue. In this view, the user can also assess the variance of feature values within cells. A larger view of a user-selected cell is shown on demand.

4.4.3 Interaction

The SOM views are interlinked with the main view so that the user can analyze the locations of the found patterns on the organ model.

– *Coloring* We use a 2D-colormap to show the distribution of various convergence patterns across the mesh (see Fig. 13). The user can directly asses where are the patterns located. Thanks to the topology-preservation property of SOMs, the user can also examine whether there are larger areas of similar types of landmark movements (areas with similar colors). As the cell position within SOM depends on its initialization (SOM can be "rotated"), the color distribution on the organ is more important than the color value in itself. We use the colormap proposed in [14].

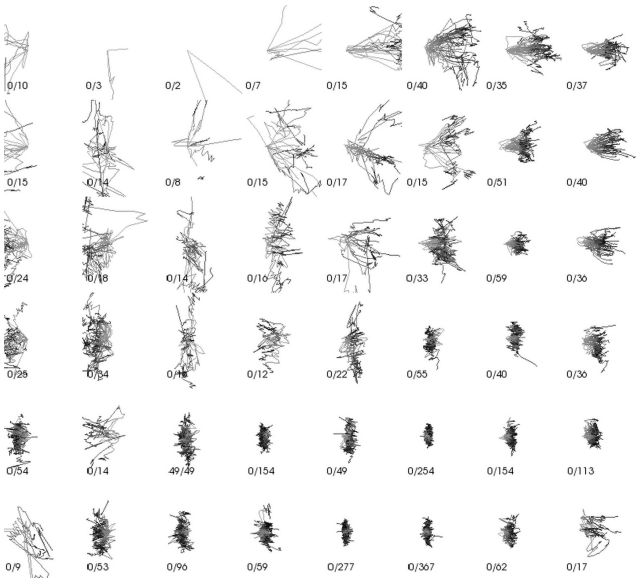

Fig. 11 SOM Trajectory view showing landmark patterns using distance to target and projection progress features. Each cell contains landmarks with similar movements.

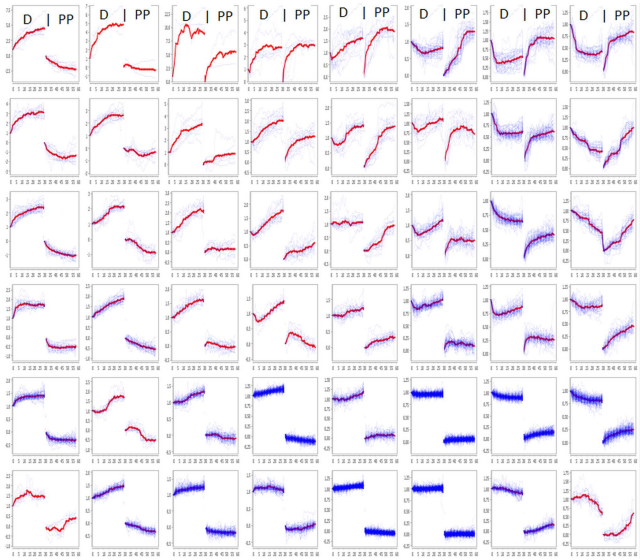

Fig. 12 SOM Feature view showing landmark patterns using distance to target (D) and projection progress (PP) features. Cells contain landmarks with similar features. Blue color: landmark features. Red color: Cell center.

– *Highlighting* The user can highlight a pattern in the SOM view and see its location on the organ. In this way, the user can assess its exact distribution across the organ (see Fig. 14 top).
The highlighting is also possible in the main view, where the user can select a particular mesh area. The selected points are then highlighted in the SOM view allowing to analyze their convergence patterns (see Fig. 14 bottom).

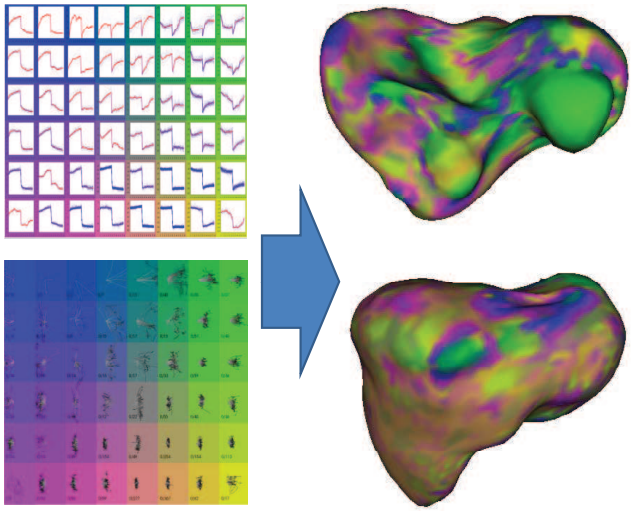

Fig. 13 The use of 2D coloring for showing the locations of landmark patterns.

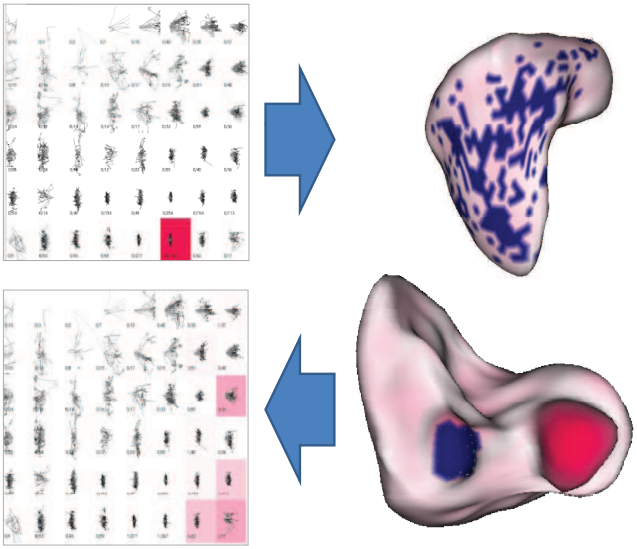

Fig. 14 Highlighting. Top: from SOM to the organ. Bottom: Landmarks movements to SOM.

5 Application Example

As a proof of concept, we apply our approach on real-world medical image data – the segmentation of a liver. The data was extracted from CT scans. The intraslice resolution ranges from 0.58 x 0.58 mm to 1 x 1 mm and the interslice resolution ranges from 0.7 to 5 mm.

We show a confirmatory analytical scenario explaining the capabilities of the rather complex system. In our case, a user selected one liver sample for a detailed inspection, because it has a cancerous tissue, unlike other samples. The user wishes to analyze how the segmentation behaves in this special case.

First, the user looks at the progress of *global convergence criteria* for an overview of the main process characteristics (see Fig. 9).

– It can be seen that average distance-based measures improve steadily during the first part of the process (Fig. 9 top), while leveling off in the second part.
– Interestingly, the Hausdorff distance (HD) shows a sharp drop between iterations 13 and 17 (Fig. 9 center). As the Hausdorff distance is a maximum of all landmarks distances, the two views would suggest that there is a part of the organ having bad quality, while the rest of the organ improves.
– A look at the stabilization metrics (Fig. 9 bottom) shows a peak in the middle and then higher stabilization. The peak corresponds to the break seen above.
– Using linked views, the user analyzes what happened during this break (see Fig. 15). She sees two interesting regions (1,2). She notices that the Region 2 improves significantly, while the Region 1 is stable.

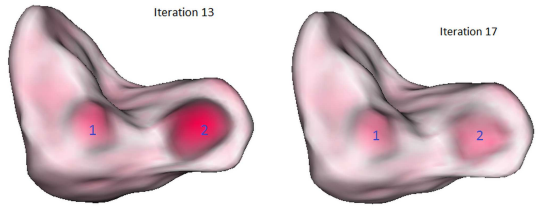

Fig. 15 View on the organ in user-selected iterations based on global convergence criteria (Fig. 9)

The user now wishes to analyze *local convergence patterns*. As a basis, the target stability is checked. Figure 6b confirms the stability, especially towards the end of the process (see Fig. 6e).
The inspection of SOM results (see Fig. 11 and 12) reveals a large variability of movement patterns.

– The top left: Movements away from the target.

- The bottom left: Small movements around the start.
- The bottom right: Movements towards the target.
- The top right: Movements to the target. We can also see the exact movement pattern: large changes at the beginning and then stabilization close to the target.

The user now employs 2D coloring to analyze how these patterns are distributed across the organ (see Fig. 13). The analysis shows several interesting insights. There is a large part of the liver in brown/orange color corresponding to regions with almost no improvement during the segmentation. There are also small regions with different movements: blue: deteriorating quality during segmentation (away from the target), green: meeting the target, and pink: increasing distance from the target but moving along the path to the target (improving PP).

The user now focuses on the two regions identified above. The complete (bad) Region 2 has the same green color corresponding to high quality improvement towards the target with stabilization in the second half of the process. She then uses highlighting (see Fig. 14 bottom) to analyze the patterns in the Region 1. The highlighted cells in the SOM are distributed in the bottom right corner – steady but insufficient improvement towards their targets (see Fig. 14).

The above findings serve as a starting point for further inspection of the algorithm by the experts. For example, the user can further analyze image features of the identified important regions and iterations (e.g., region 2 in iterations 13 and 17). Why are these important movements happening in these regions? Which image features lead to these bad regions? Which other image features would improve the results? These insights could point to possible improvements of the segmentation algorithm, which span beyond the scope of the paper.

6 Discussion

We presented a new approach for the analysis of image segmentation process.

Our method is based on a set of global and local convergence criteria. The local criteria use a single target (center target) for their calculation. We provide means analyzing the target stability, which confirm this assumption for our use case. However, there may be cases, where the targets vary strongly. In these cases, we will need to adapt the criteria used in our data analysis method.

The convergence criteria are analyzed either directly in an interactive visualization or are clustered and visualized using self-organized maps (SOM). The SOM has been chosen for its specific properties: good clustering results and direct visualization of results. This has been also shown in the use case using real-world data. Alternative clustering methods (e.g., hierarchic or density-based clustering) could be considered as well. However, they would require extensions to the employed data analysis features (e.g., a definition of density) possibly leading to a more complex analysis.

Our tool currently focuses on the analysis of the segmentation of one organ selected by the user from the training data set. This is very useful for analyzing organs with particular properties, however, there can be cases when the experts wish to compare segmentation progress in the whole data set (several data samples/organs). We would like to extend our system also for this use case. It will require a new method for identification of similar areas in several data samples and new visualizations of these results.

The advantages of our approach have been shown on a real liver segmentation data set. Nevertheless, the proposed method is generic so that it can be used also for the segmentation of other organs such as kidney, bladder or vertebra.

Our method offers experts in algorithm developments important insights into the process of image segmentation, which has been exemplified on an real-world analysis process. It would be advantageous to conduct a broader user study with multiple users in order to confirm these results and use the feedback for extending our system.

7 Conclusions

In this paper, we presented a new approach for visual analysis of SSM-based image segmentation. It offers deeper insights into the convergence behavior of the segmentation process on global (organ) and local (landmarks and regions) level. We proposed a new system with several linked views and analytical functions for this purpose. We applied it to liver segmentation data.

In the future, we would like to address the limitations stated in the Discussion section. Moreover, our approach can be extended with automatic determination of interesting views on the data (e.g., automatic positioning of the organ emphasizing the areas with bad convergence), or automatic identification of important steps in the process (e.g., iterations with a high change of segmentation quality).

Acknowledgments

The work has been partially supported by the DFG SPP 1335 project "Visual Analytics Methods for Modeling in Medical Imaging". The authors would like to thank J. Beutel for his support with the project.

References

1. Andrienko, G., Andrienko, N., Bremm, S., Schreck, T., Von Landesberger, T., Bak, P., Keim, D.: Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. Computer Graphics Forum **29**(3), 913–922 (2010)
2. Andrienko, N., Andrienko, G.: Spatial generalization and aggregation of massive movement data. Visualization and Computer Graphics, IEEE Transactions on **17**(2), 205–219 (2011)
3. Andrienko, N., Andrienko, G.: Visual analytics of movement: An overview of methods, tools and procedures. Information Visualization **12**(1), 3–24 (2013)
4. Angelelli, P., Viola, I., Nylund, K., Gilja, O.H., Hauser, H.: Guided visualization of ultrasound image sequences. In: Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM), pp. 125–132 (2010)
5. Bruckner, S., Möller, T.: Isosurface similarity maps. Computer Graphics Forum **29**(3), 773–782 (2010)
6. Busking, S., Botha, C.P., Ferrarini, L., Milles, J., Post, F.H.: Image-based rendering of intersecting surfaces for dynamic comparative visualization. Vis. Comput. **27**(5), 347–363 (2011)
7. Busking, S., Botha, C.P., Post, F.H.: Dynamic Multi-View Exploration of Shape Spaces. Computer Graphics Forum **29**(3), 973–982 (2010)
8. 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)
9. Demšar, U., Fotheringham, A.S., Charlton, M.: Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics. Information Visualization **7**(3), 181–197 (2008)
10. Dick, C., Burgkart, R., Westermann, R.: Distance visualization for interactive 3d implant planning. Visualization and Computer Graphics, IEEE Transactions on **17**(12), 2173–2182 (2011)
11. Heimann, T., van Ginneken, B., Styner, M., et al.: Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Transactions on Medical Imaging **28**, 1251–1265 (2009)
12. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis **13**(4), 543–563 (2009)
13. Heimann, T., Münzing, S., Meinzer, H.P., Wolf, I.: A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation. In: Information Processing in Medical Imaging, pp. 1–12 (2007)
14. Himberg, J.: A som based cluster visualization and its application for false coloring. In: Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on, vol. 3, pp. 587 –592 vol.3 (2000). DOI 10.1109/IJCNN.2000.861379
15. Kohonen, T.: Self-Organizing Maps, 3rd. edn. Springer, Berlin (2001)
16. Maciejewski, R., Rudolph, S., Hafen, R., Abusalah, A., Yakout, M., Ouzzani, M., Cleveland, W., Grannis, S., Ebert, D.: A visual analytics approach to understanding spatiotemporal hotspots. Visualization and Computer Graphics, IEEE Transactions on **16**(2), 205–220 (2010)
17. Matkovic, K., Gracanin, D., Jelovic, M., Ammer, A., Lez, A., Hauser, H.: Interactive visual analysis of multiple simulation runs using the simulation model view: Understanding and tuning of an electronic unit injector. Visualization and Computer Graphics, IEEE Transactions on **16**(6), 1449–1457 (2010)
18. Mayer et al.: Java SOMToolbox. online. URL `http://www.ifs.tuwien.ac.at/dm/somtoolbox/index.html`. Http://www.ifs.tuwien.ac.at/dm/somtoolbox/index.html, accessed 8.2.2013
19. Preim, B., Bartz, D.: Visualization in medicine: theory, algorithms, and applications. Morgan Kaufmann Pub (2007)
20. Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., Andrienko, G.: Visually driven analysis of movement data by progressive clustering. Information Visualization **7**(3-4), 225–239 (2008)
21. Schreck, T., Bernard, J., von Landesberger, T., Kohlhammer, J.: Visual cluster analysis of trajectory data with interactive kohonen maps. Information Visualization **8**(1), 14–29 (2009)
22. Silva, S., Madeira, J., Santos, B.: Polymeco-a polygonal mesh comparison tool. In: Information Visualisation, 2005. Proceedings. Ninth International Conference on, pp. 842–847. IEEE (2005)
23. Tversky, B., Morrison, J.B., Betrancourt, M.: Animation: can it facilitate? International Journal of Human-Computer Studies **57**, 247–262 (2002)
24. Vesanto: Som-based data visualization methods. Intelligent Data Analysis **3**(2), 111–126 (1999)
25. Willems, N., Van De Wetering, H., Van Wijk, J.J.: Visualization of vessel movements. Computer Graphics Forum **28**(3), 959–966 (2009)
26. Zhou, L., Pang, A.: Metrics and visualization tools for surface mesh comparison. Ph.D. thesis, University of California, Santa Cruz 2001. (2001)