“THEME ARTICLE”, “FEATURE ARTICLE”, or “COLUMN” goes here: The theme topic or olumn/department name goes after the colon.

Hierarchical Image Semantics using Probabilistic Path Propagations for Biomedical Research

Christina Gillmann
University of Kaiserslautern

Tobias Post
University of Kaiserslautern

Thomas Wischgoll
Wright State University

Hans Hagen
University of Kaiserslautern

Ross Maciejewski
Arizona State University

Image segmentation is an important subtask in biomedical research applications, such as estimating the position and shape of a tumor. Unfortunately, advanced image segmentation methods are not widely applied in research applications as they often miss features, such as uncertainty communication, and may lack an intuitive approach for the use of the underlying algorithm. To solve this problem, this work fuses a fuzzy and a hierarchical segmentation approach together, thus providing a flexible multi-class segmentation method based on probabilistic path propagations. By utilizing this method, analysts and physicians can map their mental model of image components and their composition to higher level objects. The probabilistic segmentation of higher order components is propagated along the user-defined hierarchy to highlight the potential of improvement resulting in each level of hierarchy by providing an intuitive representation. The effectiveness of this approach is demonstrated by evaluating our segmentations of biomedical datasets, comparing it to state of the art segmentation approaches, and an extensive user study.

INTRODUCTION

Image segmentation is typically defined as the process of dividing an image into different regions, called segments (classes), each correlating to an object in the image. This task is often required in biomedical imaging to identify specific features, such as tumors or anomalies. In order to achieve this, new methods need to be developed that can enable biomedical researchers to define novel and patient specific models and develop new diagnosis methodologies.

Several types of segmentation methods are available but, contrary to other fields, the biomedical domain defines specific requirements for segmentation algorithms to make them usable for biomedical segmentation tasks. One class of segmentation methods, semi-automatic segmentation, benefits from the biomedical specialist's input as well as automated computation. In this way, biomedical researchers can introduce their knowledge into the segmentation process by picking seed points for different segments and help the underlying algorithm output improved results.

Although these methods introduce user knowledge into the segmentation process through seeding, they are rarely used in biomedical tasks for multiple reasons[18]. One problem is the strict assignment of an image pixel to a single segment, a widely used paradigm in segmentation methods. Biomedical images often lack a crisp border between tissue types due to the partial volume effect caused by the image reconstruction process. These strict class assignments result in sharp borders between segments which makes identifying uncertain regions difficult. For those regions, the underlying segmentation algorithm may struggle to determine which segment should be assigned to an image pixel. The lack of this uncertainty information often leads to a rejection of the whole algorithm and its results as the trustworthiness of the segmentation might be questioned. This is especially impractical for biomedical research where decision making can have a huge impact on a patient's health.

Another important aspect is the human perception of objects. The human brain is trained to identify known objects and combine them to superior objects[16].This leads to an intuitive understanding of the semantic structure of objects in an image. Existing segmentation methods lack this ability to map these semantic structures. This results in usability issues for some biomedical researchers, who may lack the background knowledge of the functionality of the applied segmentation algorithm as they cannot relate their understanding of a semantic composition of an image to a given segmentation.

This work presents a segmentation algorithm that is embedded in a visual framework to be used by experts within the biomedical domain. It is designed to lead toward a broader acceptance of segmentation algorithms within the biomedical area. The definition of probabilistic image segmentation is utilized and input parameters, such as number of resulting segments, and importance of segments provided. This increases the degree of freedom for biomedical users and provides a measure of the uncertainty of a segmentation depending on the image region. In order to allow users to map their mental model of composed objects to the segmentation process, a hierarchical tree structure with arbitrary branching degree can be designed. This results in a user-defined and dataset specific hierarchical image semantic. During this process, the probabilistic segmentation of higher order components is propagated along the user-defined hierarchy to highlight potential for improvement of the resulting segmentations. The presented approach is embedded in a visualization framework that allows biomedical specialists to review their segmentation results. Here, the probabilistic segmentation approach, as well as the resulting uncertainty of the segmentation result are visualization.

The contributions of this work include:

- A flexible multi-class segmentation based on probabilistic paths;
- A probabilistic path propagation along arbitrary user-defined hierarchical image semantics, and;
- A visualization framework to explore probabilistic segmentation results of the user-defined hierarchical image semantics suitable for biomedical research.

RELATED WORK

The following section briefly summarizes the state of the art methods for biomedical image segmentation approaches and the exploration of biomedical segmentation results and summarize the resulting biomedical requirements for segmentation results.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

Biomedical Image Segmentation

A more detailed review of probabilistic segmentation techniques can be found in Naz et al.[12] . Zheng et al.[18] extended this concept, such that it can output hierarchical results. These methods work on a fuzzy c-means clustering algorithm performing an unsupervised segmentation for a predefined number of segments. Unfortunately, this type of clustering is sensitive to noise, which makes it difficult to use for biomedical datasets as segment boundaries are often noisy and ill-defined.

Another fuzzy segmentation definition can be found in Kniss et al.[8] .This method uses two-dimensional transfer functions to allow the user to select different structures in an image. Although the computation of the structures is done in a fuzzy manner, the method does not allow a hierarchical refinement of already selected structures, which is an important feature for the biomedical domain. In contrast to the approach of Kniss et al., this paper presents a fuzzy segmentation which can be utilized by the user to map their mental model of the objects perceived in the image.

Neural network approaches, such as those presented by Milletari et al.[11], are able to segment biomedical image data in a fully automatic manner. Although their use case showed promising results, neural networks need to be configured for each single setup, for example, tumor or prostate estimation would require separate neural network configurations, which is a lengthy process. Contrary to this, our proposed approach can map arbitrary hierarchies for different settings in the segmentation result.

As Beucher mentioned in his work[3], a hierarchical segmentation provides the advantage of having different semantic levels in the segmentation output. Most approaches are based on the unsupervised creation of hierarchies, as can be reviewed in the work of Zhang et. al[1]. Approaches based on partial-volume Bayesian classification [8], and improved maximum a posteriori expectation-maximization [18] determine the segmentation based on a divide and conquer strategy. Aside from the computation, these algorithms are not using the resulting hierarchy to map model relations. As such, our approach is unique in that it allows users to create a hierarchy that maps their mental model of recognized objects.

Semantic hierarchies that express a patient's organs' structures can be performed using distance fields [9], shape models [17], or atlas registration. Although the utilized hierarchies express meaningful structures in the human body, these approaches are usually limited to a specific type of organ. Here, these highly specialized algorithms might output suitable results. However, in daily clinical routines, fast adaption to new situations is a necessity. Our approach is designed to provide maximal flexibility, thus it is able to segment arbitrary organs and their substructures.

Exploration of Image Segmentation Results

Mashburn et al.[10] introduced a feedback loop into the segmentation process where the user can review the result of his determined input and either insert more seed points or delete others. Although this approach forms a feedback loop, it does not support the users in adapting their input. Our approach utilizes an uncertainty measure to form a feedback loop that helps users adapt their input.

A variety of methods to guide the user through a hierarchical segmentation process have been developed[5, 19]. The idea behind such methods is to use a pre-segmented two-dimensional histogram where users can form groups and build hierarchies. Although this allows the user to insert a hierarchy into the segmentation result, the pre-segmentation is performed automatically which can result in regions that do not map the users' understanding of the perceived objects in the image. In contrast, our proposed methodology enables the user to map an arbitrary mapping of the objects he perceives in the image thereby creating meaningful segmentation results.

The exploration of hierarchical segmentation results was targeted in Jaja et al.[6] Users can navigate through the hierarchical segmentation result and inspect it. This paper uses this knowledge and extends it to allow users to navigate and inspect hierarchical segmentation results based on the probabilistic path propagation introduced in this paper.

Balabanian et al.[2] presented a visualization methodology that is able to map hierarchical structures in biomedical data and allow the user to explore them. We use this as a starting point for representing the hierarchical segmentation outputted by the presented segmentation approach and adjust it such that it is able to express the uncertainty captured in the segmentation result.

Requirements for Segmentations in Biomedical Research

Image segmentation has a huge potential in the biomedical research area as it is a required preprocessing step in determining the volume, shape, and motion of organs[4]. The state of the art analysis shows the need for a highly flexible and hierarchical segmentation approach, the increased focus on patient-specific image reviewing methods[13] as well as an easy to use interaction with the determined segmentation. Based on the prior state of the art analysis, a careful literature review and long discussions with researchers from the biomedical research domain we have formulated a series of design requirements for biomedical image segmentation.

- R1: Biomedical image data from different imaging techniques, such as MRI and CT, must be supported by the segmentation algorithm.
- R2: Segmentation algorithms need to be able to deal with boundary insufficiencies as biomedical image data often misses crisp boundaries due to the partial volume effect[15,20].
- R3: Biomedical image segmentation results need to be able to capture patient organs and their sub-structures to be suitable for biomedical research.
- R4: A segmentation method for the biomedical research domain needs to support an efficient workflow as biomedical users need to be able to create and review segmentation without prior knowledge of the underlying segmentation algorithm.

To the best of our knowledge, there does not exist a technique that covers all of the mentioned requirements, which forms the motivation for our approach.

METHODS

The following section describes our proposed segmentation method designed for biomedical imaging. The method combines a probabilistic multi-class segmentation based on probabilistic path propagations with an arbitrary user-defined hierarchy embedded into a visualization framework to provide a review mechanism for biomedical applications.

Probabilistic Multi-class Segmentation

The standard definition of image segmentation is the partitioning of the image into different segments. This is achieved by examining each combination of voxel-segments with a depicted metric. The segment that obtains the best value based on that metric also obtains the related voxel. This results in hard borders between the segments. Unfortunately, for border voxels it is hard to decide which segment matches the metric best as they can fit equally or almost equally to different segments. In addition, a clear best fit for the metric of a segment does not necessarily mean a clear affiliation of the corresponding voxel. In the case where all segments poorly match the metric, even the best fit is not meaningful. As illustrated by these issues, the standard definition of segmentation needs to be extended.

In contrast to a fixed class assignment of each voxel, the concept of probabilistic segmentation is used in the presented method. Here, for each class, there exists a probability that each voxel belongs to this class.

In order to combine a probabilistic segmentation with the ability to allow an arbitrary user-defined number of classes, the presented approach utilizes seed points as starting points for the segmentation calculations. The user can define an arbitrary number of classes $i \in \{1 \dots n\}$. For each

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

of these classes, the user can define a set of seed points S_i. These points are voxels in the input image for which the user can clearly determine that these points belong to the selected segment.

Based on the defined seed points, the question arises: how can we best define a probabilistic segmentation for all voxels in the input image? A metric is required that determines the distance from each voxel to the closest seed point of each class. This distance can be defined by:

$$\mathrm{D_{i(x)}} := \min_{s \in S_i} d(s, x), \qquad i \in \{1, 2 \dots n\}$$

where

$$d(s_1, s_2) = \min_{C_{S_1 S_2'}} \int_{C_1}^{C_2} r(x) * C_{S_1, S_2}(x)| \, dx$$

where $C_{S_1,S_2}(x)$ is a path connecting the pixels s_1 and s_2 and r(x) is the metric defining the similarity of connected voxels. A natural choice for this metric is the gradient between the neighbored voxels. Still, arbitrary metrics can be used.

Based on the defined distance metric, the probability for each voxel to belong to a specific segment needs to be computed to obtain a probabilistic multi-class segmentation. In order to obtain this segmentation, each voxel has to be distributed to n segments. The weighted segmentation ω of a class $i \in \{1 \dots n\}$is defined as:

$$\omega_i = \frac{(m_i * d_i)^{-p}}{\sum_{j=1}^{n} (m_j * d_j)^{-p}}$$

where m_i are the distance multipliers according to each class, and p is the exponent of the segmentation. The multiplier of a class can be used to manipulate the values obtained by the distance metric. A small multiplier scales down the distance metric values whereas a larger distance multiplier increases the values. With this parameter, the user can manipulate the importance of a class. The parameter p is the exponent controlling the transition between different classes. An exponent close to 1 results in a shallow transition between the classes whereas the transitions turn sharper as the exponent becomes larger. The user can utilize this parameter to control the blur of class borders. The probabilistic multi-class segmentation distributes the entire voxel to all segments. In the case where one class obtains a high weight, the voxel strongly belongs to that class. During the segmentation, special cases where one or more segment obtains a weighted segmentation of 0 occur. In this cases, the remaining segments share the voxel equally. This ensures that the weighted segmentation of a voxel sums up to 1 for all classes.

To allow users an intuitive and fast review of the segmentation result that cannot be assigned to a specific class, the presented probabilistic multi-class segmentation is embedded into a visual framework. Based on the probabilistic multi-class segmentation, a segmentation visualization $\hat{c}_n$ and an uncertainty visualization $\hat{u}_n$ can be derived as:

$$\hat{c}_n = \sum_{i=1}^{n} c * \omega_i \ , \alpha_i = \max(\omega_i)$$

$$\widehat{u_n} = c_u \ , a_u = \min(1 - \omega_i)$$

This results in a bipartite visualization for the probabilistic segmentation result. The segmentation visualization marks a voxel with the color of a segment if it strongly belongs to this class. For voxels that cannot be distributed into a segment, the voxel's transparency is increased. In case of equal weight distributions (when one class is zero), the present color is a mixture of the colors from all remaining classes. If a specific type of tissue is more likely than another, their colors will be blended according to their classes' weight. In contrast, the uncertainty visualization highlights voxels that cannot be assigned clearly to a segment with a highlight color.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

Figure 1(top) is the image of a CT scan slice showing the upper legs of a patient (a) and its segmentation comparing the fixed multi-class segmentation and the probabilistic multi-class segmentation. The input image is segmented into background (white), soft tissue (green), muscles (red) and bones (blue). Although a bone contains a wide range of grey scale values, they can be captured in one class in the case seed points are spanned throughout the greyscale range contained in the bone. Both segmentation results are based on the same seed points depicted by the user. Figure 1 (b) presents the fixed multi-class segmentation. In this examples multiple missclassifications occurred as indicated by the red circles. In contrast, the segmentation result achieved with our approach highlights uncertain areas in the image assigning a high alpha-value to these areas (c) and coloring uncertain areas in a highlight color (d). In contrast to the fixed class assignment in (c), the visualization presented in this paper indicates where users need to refine their input in order to achieve a proper segmentation result. As this example shows, the uncertainty between three classes is visualized: bones, fat, and muscles. Although the algorithm might output that a voxel does not belong to the bone class, the uncertainty response can be strong if the algorithm is not able to determine whether the voxel belongs to the muscle or the fat tissue of the patient.

The probabilistic segmentation design combined with the uncertainty-aware feedback loop helps users in mapping their mental model of different objects that are visible in an image. The weighted segmentation is designed such that the classes can be mapped in a flexible manner through the exponent and segment multiplier parameters.

Hierarchical Image Semantics

As human perception is based on recognizing objects and composing them into superior objects, this paper aims at assisting biomedical users in mapping their individual mental associations of patient organs and their composition to functional units. This section will introduce the concept of hierarchical image semantics that is based on a probabilistic path propagation along the user defined hierarchy.

As a consequence, a mechanism is required that can separate a segment into further subsegments. Considering the multi-class segmentation, it is possible to interpret the separation of the input image into several classes as a tree where each node represents a segment. The root node segment contains the entire input image. In the case of the multi-class segmentation, the root node obtains all resulting segments as its child segments. All segments resulting from the multi-class segmentation form the leaf nodes of the resulting tree. In order to achieve a further segmentation of any leaf node, the probabilistic segmentation concept needs to be further extended. To achieve this goal, this paper provides a segmentation structure where each leaf node i is a possible candidate for further segmentation into n child nodes. Consistently, these segmentation nodes $j \in (1 \dots n), i \notin \{1 \dots n\}$ need to assemble to the parent node n.

Here, the user can choose a segmentation node and define the number of child nodes he or she intends to separate the parent into (n). To consistently provide a probabilistic segmentation, the weight $\omega_i$of a segment is divided to the defined child segments. To restrict the propagation of the child segments outside the parent segment, the metric parameters that are used to calculate the weighted segmentation need to be adjusted according to the superior weighted segmentation. Therefore, the distance metric of a child node can be adjusted by using the weight of the parent node. Let d_{xy}be the distance between two voxels. The adjusted distance for child nodes changes to:

$$\bar{d}(x,y) = \begin{cases} \infty, & if\ w_x = 0 \\ \min\limits_{C_1, C_2} \int_{S_1}^{S_2} |m(x) * C_{S_1,S_2}(x) * \overline{w_x}| dx, & else \end{cases}$$

This results in a scaling of the metric according to the weight of the parent node. With respect to the adjusted metric $\bar{d}_i$ it is possible to distribute the weight of a parent node i to an arbitrary number of child nodes$\{1 \dots n\}$. Resulting from this, the weight of a child node j can be calculated as:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

$$\bar{\omega}_i = \frac{(m_i * \bar{d}_i)^{-p}}{\sum_{j=1}^{n}(m_j * \bar{d}_k)^{-p}}$$

The result of this calculation is a value between 0 and 1 describing the ratio of the parent's weighted segmentation that is distributed to a depicted child node j. Consistently, the child node's weight for their parent node can be summed up to one for each voxel. This does not describe the probability of a voxel to belong to a depicted segmentation node in the tree structure. The final weighted segmentation $\omega_i$for an arbitrary segmentation node can be calculated as the multiplication of all $\bar{\omega}_i$ that belong to the nodes on the path between the node i and the root node.

In order to provide the users with a tool to review their segmentation achieved by a hierachrchial image semantic, we provide intuitive visualizations. The resulting segmentation visualization $\hat{s}_i$is a recursively defined function that blends the user defined colors of the nodes of the hierarchical image semantic according to their weight. Starting from the root node, the segmentation visualization can be obtained by recursively considering the colors of the child nodes j by:

$$\hat{s_i} = \begin{cases} c_i, \ \ if\ i\ is\ a\ leaf \\ (1-\delta) * c_i + \delta * \sum_{j=1}^{n} \hat{s_j} * \omega_j, \ \ else\ (where\ \delta = max(\omega_j) \end{cases}$$

This results in a visualization that starts from the leaf nodes in the tree accumulating the colors of a node in the same manner as shown for the multi-class segmentation. After that, the resulting color is blended with the parent node color according to the maximum weight δ of the child nodes. In the case where δ is close to one, this means, that the parents weight was clearly distributed to one of the child nodes. This results in the color of this corresponding child node for the concerned voxel. This correlates to the natural understanding that the separated parts replace the composed object. In contrast to that, if the maximum weight of the child nodes is low, the resulting color is blended with the color of the parent node resulting in areas where the parent node color appears. This visual uncertainty encoding refers to the understanding that the parent object cannot be clearly separated to the child objects.

In addition, the α −value of the segmentation visualization can be calculated by multiplying the maximum weight of each layer in the tree with the maximum weight of the layer above until the root node is reached. This works as a stencil from top nodes that cut out its own uncertainty in the visualization for the underlying subtree.

Consistent with the prior visualizations, the hierarchical uncertainty visualization displays the depicted uncertainty color c_u, that is blended with an α − value. This value is defined as a recursive function multiplying the minimum weight of the child nodes with the weight of the current examined node:

$$u_{u_i} = \begin{cases} 1-\omega_i, \ \ if\ i\ is\ a\ leaf \\ (1-\omega_i) * \min(\hat{u_j}), \ \ else \end{cases}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

In the case that the uncertainty is high on a certain level in the tree, the corresponding voxels are highlighted in the visualization. The uncertainty value is used as a stencil for the underlying layers in the tree.

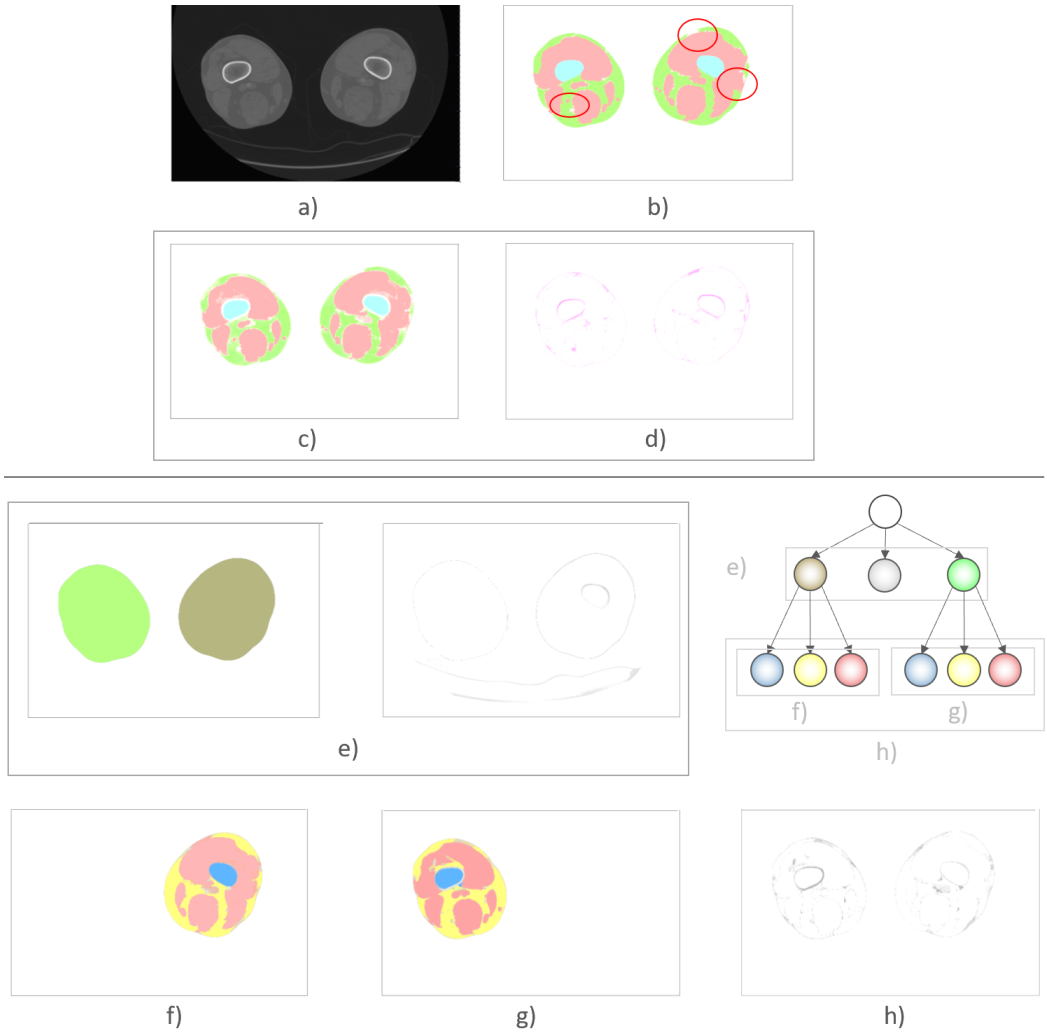

Figure 1 Top: Multi-class segmentation of a CT scan slice showing the upper legs of a patient segmented into background (white), soft tissue (green), muscles (red) and bones (blue). a) Original Image b) Fixed class segmentation containing several miss-classifications, that are highlighted with red circles. c) Probabilistic multi-class segmentation visualization indicating areas, that cannot be distributed to a specific segment through a high transparency. d) Uncertainty visualization highlighting areas that cannot be distributed to a specific segment with a highlight color. Bottom: Hierarchical Image Semantics using probabilistic path propagations applied to the example of Bottom: e) First level segmentation of the input image into background (blended out), left leg (brown) and right leg (green). f) Second level segmentation of the left leg into soft tissue (yellow), muscles (red) and bones (blue). g) Second level segmentation of right leg into soft tissue (yellow), muscles (red) and bones (blue). h) Uncertainty visualization for the second layer of the segmentation tree and a tree representation of hierarchical image semantic.

For an improved visual exploration, the user can enable or disable arbitrary nodes or entire subtrees of the segmentation structure to focus on specific segments or examine the relation between them. Figure 1 (bottom) shows how the presented method can be applied to the CT of the human leg. The resulting hierarchical image semantic is shown in the image. As the hierarchical design is not limited, different users can create different hierarchies. The visual representations consist

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

of a color-coding indicating the probabilistic segmentation result as well as an uncertainty visualization of enabled nodes. Subfigures 1 ((e),(f),(g)) show how the user can explore the segmentation result. Different nodes can be enabled and disabled resulting for the inspection of specific classes or groups of classes and their inter-class uncertainty. Users can review their segmentation result and correct the seed input of specific nodes or refine a leaf node of the tree into further nodes.

In summary, the presented approach builds a visual framework that is able to review user-defined hierarchical image semantics. The hierarchical image semantic can be designed individually by the user to encode his or her mental model of objects and their composition. In addition, the underlying segmentation algorithm is probabilistic and the resulting probability values are propagated along the user defined hierarchy to encode uncertainty information through the entire segmentation design process.

RESULTS

To show the effectiveness of the presented approach, we applied the presented segmentation to a variety of biomedical image data, performed a comparison with state of the art segmentation tools, and performed a user evaluation.

Segmentation Result for a Brain Tumor Dataset

Used libraries
VTK http://www.vtk.org/
ITK https://itk.org/
Qt https://www.qt.io/

Our approach was implemented using C++ with the vtk, itk, and Qt libraries. A geodesic distance metric is used in the presented examples to demonstrate the approach. The distance metric utilized for the presented results is the difference between neighboring voxel gradients.

Figure 2, shows our segmentation result applied to a dataset from the cancer imaging archive. The example shows a brain MRI scan that contains a tumor, and the data size is 256x215. The image is anonymous, and the patient's face has been removed. In the original image, it can be observed that the border of the tumor can not be clearly identified; however, defining the tumor boundary is critical for biomedical researchers to develop a suitable treatment plan. Although the tumor location itself could be achieved by a segmentation approach that separates the image into an inside and outside object, the remaining segments are an important information for biomedical doctors to plan surgeries and identify tissues that would be affected during the procedure.

Visit http://www.cancerimagingarchive.net/ *to learn more about the data acquisition process*

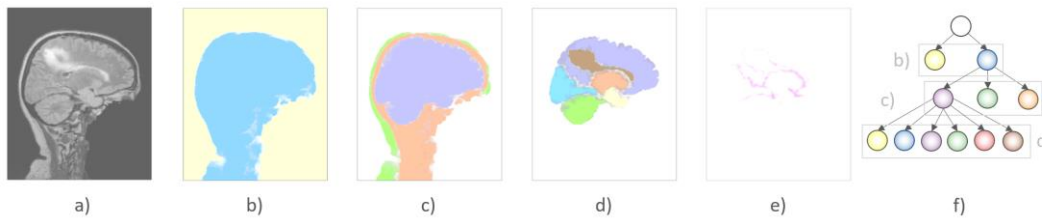

Figure 2: Hierarchical image semantics applied to a MRI scan of the brain. a) Original Image. b) First layer segmentation containing background (yellow) and foreground (blue). c) Second layer segmentation of the foreground into scalp (green), brain (purple) and other soft tissues. d) Third layer segmentation of the brain into spinal pons (yellow), cerebellum (yellow), talamus (red), lobo occipitale (blue), lobo paritale/frontale (purple) and tumor (brown). e) Uncertainty visualization of third layer segmentation. f) Tree representation of hierarchical image semantic.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

Figure 2 shows our segmentation result. The original image (Figure 2 (a)) is initially segmented into the background (yellow) and foreground (blue). After this, the foreground is further segmented into scalp (green), brain (purple) and other soft tissues. In the third segmentation step, the brain is segmented into spinal pons (orange), cerebellum (yellow), talamus (red), lobo occipitale (blue), lobo paritale/frontale (purple), and tumor (brown). The resulting hierarchical semantic is shown in Figure 2 (f). As previously mentioned, the identification of the tumor size is an important factor in clinical daily routine. This dataset represents a common scenario where a clear determination of the tumor's border is difficult using traditional segmentation methods. Our approach provides not only a segmentation but also details on the uncertainty to provide insight into the probability of whether or not a specific pixel contains tumor tissue.

Comparison with State of the Art Segmentation Tools

To demonstrate the benefits of our approach, we compared the segmentation result of the presented approach with open source segmentation tools.

Compared open source tools
ITK https://itk.org/
ITK Snapl http://www.itksnap.org/pmwiki/pmwiki.php
Slicer https://www.slicer.org/
ImageJ/FIJI https://imagej.nih.gov/ij/index.html
MITK http://mitk.org
Free trial of MIPAR http://www.mipar.us/

Comparison of the ability to fulfill the defined requirements

For each of the tools, we identified the different possibilities to perform segmentation tasks and applied each segmentation algorithm to the leg example. Here, the task was to separate the muscles of the patient's left leg from the rest of the image as shown in the groundtruth (Figure 3 (k)). The groundtruth is a manual segmentation performed by our biomedical collaborators.

The goundthruth and all generated results are shown in the same color scale. White pixels mean, the underlying algorithm detected muscle tissue, whereas black pixels mean, that the algorithms did not detect muscle tissue. When applicable, the probability of an image pixel to show muscle tissue is encoded by a color scale reaching from black (probability = 0) to white (probability = 1).

ITK Library Figure 3 shows the segmentation results generated by the ITK implementation using a region growing approach (a), watershed approach (b) and an active contour approach (c). In general, ITK requires programming skills (R4) and does not support user interaction right away.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

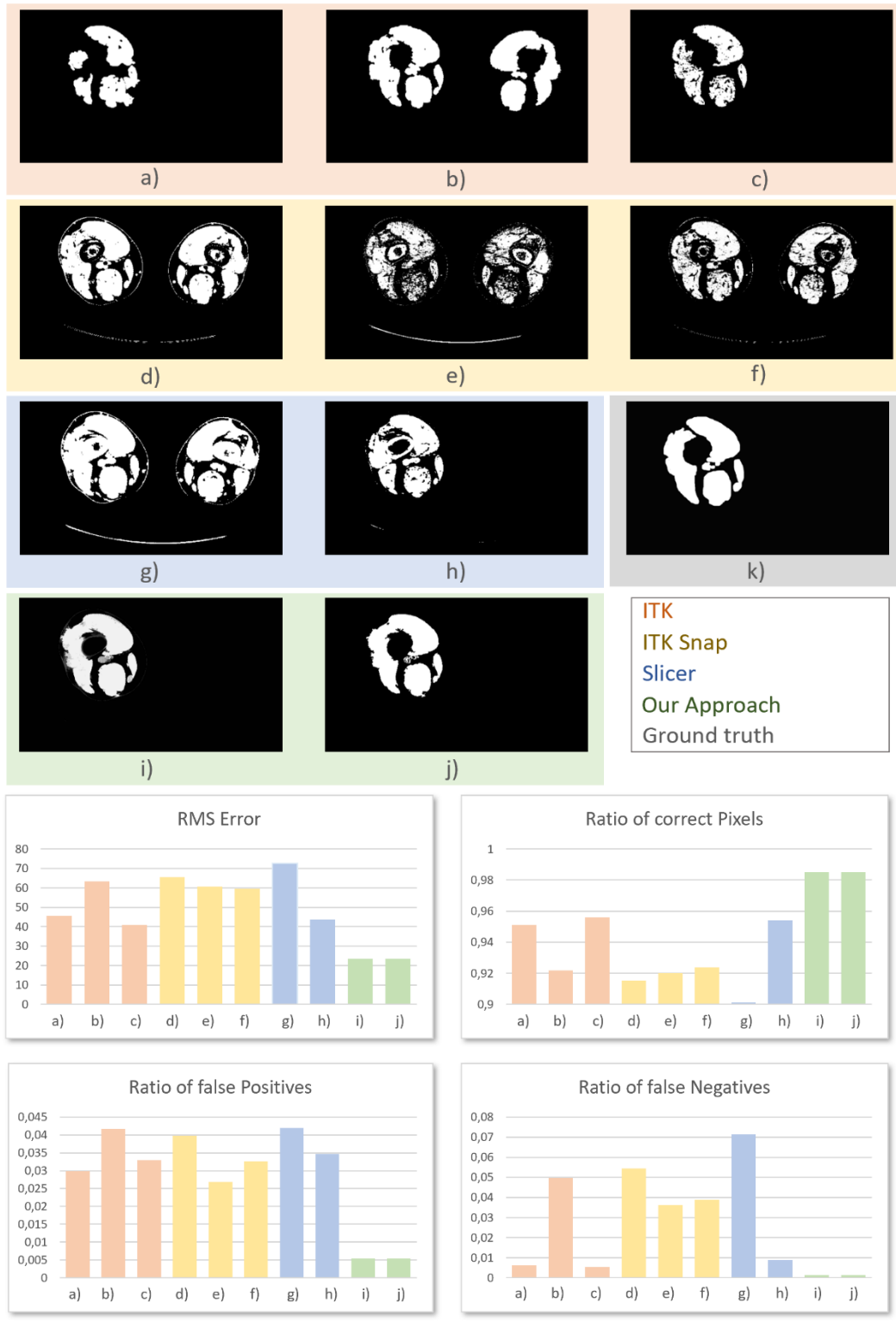

Figure 3, top: Comparison of the presented method for the example shown in Figure 1 with state of the art methods provided by the ITK library (a-c), ITK Snap (d-f), 3D Slicer (g-h), and the presented approach (i-j) utilizing a groundtruth. a) Connected threshold segmentation result. b) Watershed segmentation result. c) Geodesic contour level set segmentation result. d) Region growing segmentation result. e) k-means clustering segmentation result. f) Thresholding segmentation results. g) Active Contour Segmentation result. g) Region growing segmentation result. i) Presented

approach. j) Presented approach with a thresholding of all pixels that have a lower prob-ability then 95%. Bottom: Statistical analysis of the compared segmentation results utilizing the groundtruth. Top left: root mean square error of all pixels. Top Right: ratio of correctly identified pixels. Bottom left: ratio of false positive pixels. Bottom right: ratio of false negative pixels.

The segmentation result for the connected threshold segmentation approach is shown in Figure 3 (a). Although this approach seems promising, it contains holes in the segmented muscles which are not present in the original image as shown in the groundtruth segmentation.

Figure 3 (b) shows the result of the watershed segmentation approach. In general, the watershed approach separated the image into different levels. Here, a further step to select the levels that encode muscle tissue was conducted. It can be seen that the watershed algorithm is not able to separate the muscle tissue from the left and right leg.

Figure 3 (c) shows the segmentation result generated by the active contour segmentation approach. Although the algorithm does contain fewer holes than the connected threshold segmentation approach, it is not able to identify all pixels of the image that show muscle tissue.

ITK Snap is an open source tool. It provides a variety of image processing tools including segmentation algorithms. We tested the available segmentation algorithms covering region growing, c-means clustering and thresholding.

Figure 3 (d) shows the result for a region growing based segmentation approach. Although the seeding can be achieved in an intuitive manner (R4), the algorithm was not able to separate the muscles from the left and right leg and contains a lot of missclassifications.

Figure 3 (e) and (f) show the results for a k-means clustering and a threshold based clustering. ITK Snap provides an easy to use adaption of the input parameters (R4), such as number of k and the thresholds, but the results are not sufficient due to a variety of misclassifications.

3D Slicer is an open source tool, that works out of the box similar to ITK Snap. It provides two different segmentation approaches. Figure 3 (g) and (h) show the results for a thresholding approach and a region growing approach.

Our Approach Figure 3 (i) and (j) show the segmentation results of the presented approach. Figure 3 (i) shows the probabilistic results whereas Figure 3 (j) shows a thresholded result presenting all pixels that contain muscle tissue with a probability of minimum 95 percent. The result shows that the presented methods outputs the muscle of the right foot correctly. In case of partial volume effects in the input image, the algorithm is able to express the resulting uncertainty (R1). Here, the user can decide to adjust their input parameters (R3) as the method is embedded into a visual framework. In addition, the presented approach holds the flexibility to re-segment each of muscles and therefore introduce the semantic understanding of object composition into the segmentation result.

We investigated the ability to address the formulated requirements of further open source segmentation tools including ImageJ/FIJI, MITK and the free trial of MIPAR. The results of all tested tools show, that these tools lack the ability to handle boundary insufficiencies (R2) and express sub-structures of segmentation nodes in their segmentation result (R3). From all tested approaches, the only method that is able to capture all mentioned requirements is the presented approach.

Statistical comparison of the segmentation results to achieve the groundtruth result

In order to obtain an objective comparison, we applied a variety of measures to rate method correctness. First, we utilized the root mean square error of all image pixels when directly comparing all image pixels of a segmentation result with the groundtruth. Second, we computed a ratio of correctly determined voxels. Finally, we determined the ratio of false positives and false negatives. The results can be found in Figure 3 (bottom).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

The results for the root mean square error shows that our proposed approach has the smallest error in comparison to all other tested segmentation approaches, Figure 3 (bottom). Furthermore, our approach had the highest accuracy (98,5%), which is at least 4% more than the other segmentation approaches we compared to.

When considering the false positives of the tested segmentation results, our approach achieves a ratio of 0.005 (0.5%). The second best approach (e)) has a ~2.7% false positives. Finally, when considering the ratio of false negatives, our approach again achieves the best results (0.001) in comparison to the remaining segmentation approaches.

The results clearly show, that the accuracy is increased by our proposed method in both settings (probabilistic and non-probabilistic), while addressing the needs of biomedical researchers.

User Study

In order to show the effectiveness of our approach, a user evaluation was performed. We used a questionnaire-based approach where participants were asked to rate segmentation results according to different aspects. First, computer science PhD students with and without prior knowledge of segmentation algorithms were asked to participate. Next, human biology researchers that are using segmentation algorithms during their daily tasks participated in the evaluation.

In the questionnaire, the participants were asked to review an image of a very simple chair and describe it as adequately as possible. The reason for using this chair model was that it does not require any domain knowledge since everyone typically is familiar with a chair. All participants mentioned the chair and its visible components (4 legs, seat shell, and felt slides). The descriptions made by the participants confirm the theory of human perception where superior objects are composed of sub-objects. After that, they had to review a segmentation result that was visually encoded in four different ways: (1) fixed class-assignment and multi-class segmentation; (2) probabilistic class assignment and multi-class segmentation; (3) fixed class assignment and hierarchical segmentation, and; (4) probabilistic class assignment and hierarchical segmentation (our approach). The participants were asked to rate two aspects of the segmentation visualizations: the ability of each technique to show segmentation errors (Q1), and; the ability of each technique to map to their personal understanding of the object's structure (Q2). The possible scores followed a Likert scale ranging from 1 (not at all) to 5 (perfectly).

Evaluation with Computer Science PhD Students

In total 24 people participated: 18 men and 6 women. They were asked whether they have issues with color-blindness. Only one participant answered with yes, but mentioned it was solely a slight red-green blindness. The participants' ages where between 24 and 35.

The results can be seen in Figure 4 (top). The results for question 1 can be found in the upper image where the first four graphs show the resulting rating of the four presented techniques. The user experience with segmentation techniques is color coded, where blue means experienced users and red means inexperienced users. Overall, our method was rated best for the ability of each technique to show segmentation errors. Results indicated that experience did not impact the rating. We also observed that inexperienced users tend to give higher ratings, as observed in the last graph showing the average of the ratings.

The results for the rating of ability of each technique to map the personal understanding of the objects structure is shown in the lower part of Figure 4 (top). In the overall rating, method 3 was rated best when basing that determination on the average alone. However, there were some participants that disliked method 3 whereas there were no participants in those categories for our method. Most users reported being distracted by the probabilistic segmentation visualization when concentrating on the hierarchical aspect of the segmentation result. This can be explained by the fact that the human eye is trained to identify sharp borders. Still, our method is rated very well with an average of 4 points. The standard deviation is the lowest of all four methods for question 2 indicating a very consistent perception of our approach among the participants. When reviewing the difference between experienced and inexperienced users, it can be seen that expe-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MCG.2019.2894094, IEEE Computer Graphics and Applications

rienced users rated our approach almost as highly as method 3. But again, even among the experienced users, there were some that disliked method 3. This may indicate that experienced users are aware of the benefits of probabilistic segmentation visualizations. Hence, a short training phase for our method may be beneficial to enable its full potential. Overall, however, our method was accepted at least as well as method 3 considering the negative votes on method 3.

Figure 4: User evaluation of four different segmentation techniques. Method 1: fixed multi-class segmentation. Method 2: probabilistic multi-class segmentation. Method 3: fixed hierarchical segmentation. Method 4: Our Method. The users reviewed a segmentation containing miss-classifications performed with the four approaches. Then the users have been asked to rate the ability of each technique to represent segmentation errors (top) as well as the ability of each technique to map object parts and their composition (bottom). The values for rating on a Likert scale ranged from 1 (not at all) to 5 (perfect). The graphs show the rating of each method by the user where experienced users are colored in blue and not experienced users in red. The average rating is plotted in the last graph for each technique.

Evaluation with Human Biology Researchers

As Figure 4 (bottom) shows, the user acceptance for the presented method is higher when users already know the concept of segmentation and have previous experience working with those techniques. We examined this finding further by performing a second stage of the evaluation considering domain scientists from the human biology area. They are used to segmentation algorithms and work with them on a regular basis.

In total, 8 domain scientists participated with their age ranging from 23 to 42. The participants are PhD students and staff members of a human biology department. During their daily tasks, these biologists work with a variety of biomedical image data such as CT and microscopic images. The priming evaluation example was the same as for the computer science PhD students.

The results of the evaluation can be reviewed in Figure 4 (bottom). Our method was rated best for the ability to indicate segmentation errors as well as the ability to map the user's understanding of the object's structure. In addition, the standard deviation of our method is, in both cases, very small. These results show that our method was consistently preferred by users from the biomedical domain.

To show the statistical significance of the outcome of the user study, we performed a series of t-tests. We tested the statistical significance of the presented approach in comparison to the remaining segmentation results from the user evaluation. We performed a one-sided t-test with similar standard deviation using the rates from the biomedical experts. As we are targeting this group, we are mostly interested in the statistical significance of their ratings.

When assuming that our method is not significantly better than each of the methods we targeted, we can check this hypothesis with the t-test by using a significance value of 5%. Method 1 (results: Q1: 0.014 and Q2: 0.012) and method 2 (results: Q1: 0.048 and Q2: 0.026), was rated significantly better. For method 3 (results: Q1: 0.408 and Q2: 0.149), our method is not rated significantly better. Still, the overall rate for the presented method is higher than the rating for method 3. As such, our user studies indicate that our approach is better received among the study participants.

We also presented our visualization to users from the biomedical research domain and they were very satisfied with the interaction methodologies and the time required to create a segmentation. While we did not perform specific measurements on the time of use, the users typically required ~1 minute to achieve an initial segmentation and additional 5-10 minutes to refine their results. This time consumptions are equal for all tested segmentation approaches.

These results are strengthened by the positive feedback of domain scientists when we demonstrated our method applied to their datasets. The most promising are listed below:

> *"With our current methods we are not able to express uncertainty, and this is a feature we really appreciate."*
>
> *"My self-defined hierarchy helped me understand the segmentation result."*
>
> *"If this method was available for us earlier, we would have been able to produce our results faster."*
>
> *"I worked with a couple of segmentation approaches before and I was very disappointed, that none of them allowed a visual inspection such as the presented approach"*

DISCUSSION

The presented approach is highly flexible allowing different biomedical image data, such as CT and MRI (R1). Furthermore, the underlying concept should be applicable to datasets that are defined on a regular grid. Depending on the underlying image data, users can adjust the utilized distance metric. If the metric outputs strongly incorrect results for a voxel to belong to a specific class, the presented approach is not able to identify this. For biomedical image segmentation, the

geodesic distance outputs suitable results. For specific cases, such as vessel segmentation, adapted distance metrics based on the vesselness measure can be utilized.

The result of the presented approach is a segmentation containing user defined segments that shows uncertain regions between different segments (R2). These regions occur between segments where the underlying approach cannot clearly determine which segment a voxel belongs to. This helps users to refine their input, especially the selection of seed points, and therefore refine their segmentation result.

Users can choose an arbitrary number of resulting segments and seed points, segment magnification and the type of transition between these segments, and the type of distance metric used (R3). The presented probabilistic multi-class segmentation is embedded in a hierarchical framework where users can re-segment arbitrary segments. This results in a hierarchy and, as this hierarchy is not created algorithmically, it encodes the user's understanding of perceived objects and their composition as related to more complex objects. The presented results show that this is identical to the user's way of describing objects in an image. It resembles a hierarchical image semantic that encodes a user's mental model of their data. The design of this semantic is very flexible as there are no restrictions as to which nodes can be re-segmented or in how many classes each layer on the tree is allowed to be contained. In cases where the resulting tree structure is known (such as shown in the example of the foot segmentation) the tree could be handed to users and the only part left for them to do is to fill it with suitable parameters. In addition to the already available parameters, the system could be augmented to allow the use of different metrics for different segmentation nodes. This could be helpful in scenarios where specific structures can be described through a metric.

The approach of hierarchical image semantics is embedded into a visual framework, where users can interactively explore their segmentation results and refine it if necessary (R4). In this framework, a two-part visualization (segmentation result visualization and uncertainty visualization) is provided. In addition, the hierarchical image semantic is visualized as a tree that can be used for navigation by the user. The user can assign different colors to each node in the hierarchical image semantic. Furthermore, users can enable or disable arbitrary nodes or subtrees in the hierarchical image semantic visualization. Based on this selection, the segmentation result visualization and the uncertainty visualization are updated, which adapts the visualization. Although the presented approach requires a training period for the user to become familiar, they rated our approach as very helpful in expressing the uncertainty of a segmentation result and encoding their mental modal.

CONCLUSION

Our framework combines the abilities of a probabilistic segmentation approach and a hierarchical segmentation approach where users can determine the number of segments, their seed points, and further adjust parameters to improve the image segmentation results in biomedical applications. The segmentation algorithm is based on a geodesic path approach where arbitrary metrics can be used to achieve a segmentation result. In order to map the user's individual recognition, any segment can be re-segmented into more fine-grained components. This re-segmentation is achieved by propagating the probability of each voxel belonging to the selected segment along with the user-defined hierarchy called hierarchical image semantic. This allows an examination of areas in the segmentation, where the segmentation is uncertain. The presented approach is embedded in a visual framework where users can explore their segmentation through intuitive interaction with the defined hierarchy.

The effectiveness of this approach was demonstrated for different biomedical datasets containing complex structures. In addition, we performed a user study showing with users from the human biology field that the combination of probabilistic and hierarchical segmentation contributed to an improved user experience. Al last, the effectiveness of the presented approach was demonstrated through an extensive comparison with state of the art approaches.

REFERENCES

1. Image segmentation evaluation: A survey of unsupervised methods. Computer Vision and Image Understanding, 110(2):260 – 280, 2008.
2. J.-P. Balabanian, I. Viola, and E. Groller. Interactive illustrative visualization of hierarchical volume data. In Proceedings of Graphics Interface 2010, GI '10, pages 137–144, Toronto, Ont., Canada, Canada, 2010. Canadian Information Processing Society.
3. S. Beucher. Geodesic Reconstruction, Saddle Zones & Hierarchical Segmentation. 2001.
4. M. Gao, J. Huang, X. Huang, S. Zhang, and D. Metaxas. Simplified labeling process for medical image segmentation. In Medical Image Computing and Computer-Assisted Intervention, volume 7511 of Lecture Notes in Computer Science, pages 387–394. Springer Berlin Heidelberg, 2012.
5. C. Y. Ip, A. Varshney, and J. JaJa. Hierarchical exploration of volumes using multilevel segmentation of the intensity-gradient histograms. IEEE Transactions on Visualization and Computer Graphics, 18(12):2355– 2363, 2012.
6. J. JaJa, A. Varshney, and C. Y. Ip. Hierarchical exploration of volumes using multilevel segmentation of the intensity-gradient histograms. IEEE Transactions on Visualization & Computer Graphics, 18:2355– 2363, 2012.
7. J. M. Kniss, R. L. V. U. Jr., A. Stephens, G.-S. Li, T. Tasdizen, and C. D. Hansen. Statistically quantitative volume visualization. In IEEE Visualization, pages 287–294. IEEE Computer Society, 2005.
8. D. H. Laidlaw, K. W. Fleischer, and A. H. Barr. Partial-volume bayesian classification of material mixtures in mr volume data using voxel histograms. IEEE Transactions on Medical Imaging, 17(1):74–86, Feb 1998.
9. G. E. Marai, D. H. Laidlaw, and J. J. Crisco. Super-resolution registration using tissue-classified distance fields. IEEE Transactions on Medical Imaging, 25(2):177–187, Feb 2006.
10. Mashburn, Lynch, Ma, and Hutson. Enabling user-guided segmentation and tracking of surface-labeled cells in time-lapse image sets of living tissues. PubMed, 5, 2012.
11. F. Milletari, N. Navab, and S. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. CoRR, abs/1606.04797, 2016.
12. S. Naz, H. Majeed, and H. Irshad. Image segmentation using fuzzy clustering: A survey. In Emerging Technologies (ICET), 2010 6th International Conference on, pages 181–186, Oct 2010.
13. E. S. of Radiology (ESR). Medical imaging in personalised medicine: a white paper of the research committee of the european society of radiology (esr). Insights into Imaging, 6(2):141–155, Apr 2015.
14. J. R. Parker. Algorithms for Image Processing and Computer Vision. John Wiley & Sons, Inc., 1st edition, 1996.
15. S. Wang, L. Li, H. Cohen, S. Mankes, J. J. Chen, and Z. Liang. An em approach to map solution of segmenting tissue mixture percentages with application to ct-based virtual colonoscopy. Medical Physics, 35(12):5787– 5798, 2008.
16. J. Weber. The Judgement of the eye : the metamorphoses of geometry - one of the source of visual perception and consciousness (a further development of Gestalt psychology). Springer, Vienne, New York, 2002.
17. F. Yokota, T. Okada, M. Takao, N. Sugano, Y. Tada, N. Tomiyama, and Y. Sato. Automated CT Segmentation of Diseased Hip Using Hierarchical and Conditional

Statistical Shape Models, pages 190–197. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

18. Y. Zheng, B. Jeon, D. Xu, Q. J. Wu, and H. Zhang. Image segmentation by generalized hierarchical fuzzy c-means algorithm. J. Intell. Fuzzy Syst., 28(2):961–973, Mar. 2015.
19. Christina Gillmann, Robin G. C. Maack, Tobias Post, Thomas Wischgoll and Hans Hagen. An Uncertainty-aware Workflow for Keyhole Surgery Planning using Hierarchical Image Semantics, Visual Informatics, 2(1):26-36, Mar. 2018
20. Gillmann, C., Wischgoll, T., & Hagen, H.. Uncertainty-Awareness in Open Source Visualization Solutions. IEEE VIS, Vis in Practice, (2016).

ABOUT THE AUTHORS

Christina Gillmann earned her PhD in Computer Science at the University of Kaiserslautern in 2015. She is part of the Computer Graphics and HCI group and a member of the International Research Training Group 2057 on Physical Modeling for Virtual Manufacturing Systems and Processes. She is working in the area of image processing, especially under the aspect of uncertainty.

Tobias Post earned his PhD at the Computer Graphics and HCI Group of the University of Kaiserslautern and is a member of the International Research Training Group (IRTG) 2057 on Physical Modeling for Virtual Manufacturing Systems and Processes. His research interests are visual analysis, parameter space exploration and comparative/ensemble visualization.

Thomas Wischgoll is currently a full professor and the Director of Visualization Research at Wright State University. His research interests include scientific visualization, flow and scientific visualization, data analytics, virtual environments and display technology, as well as biomedical imaging and visualization.

Hans Hagen is head of the Computer Graphics \& HCI Group at the University of Kaiserslautern. His research interests include geometric modeling, scientific visualization,computer graphics, and information visualization. Hagen has a PhD in mathematics from the University of Dortmund. He has received several honors including three IEEE awards.

Ross Maciejewski is an Associate Professor in the School of Computing, Informatics and Decision Systems Engineering at Arizona State University. His primary research interests are in the areas of geographical visualization and visual analytics focusing on public health, dietary analysis, social media, and criminal incident reports.