Exploring Time Series Segmentations Using Uncertainty and Focus+Context Techniques

C. Bors[1], C. Eichner[2], S. Miksch[1], C. Tominski[2], H. Schumann[2], and T. Gschwandtner[1]

[1]Institute of Visual Computing and Human-Centered Technology, TU Wien, Austria
[2]Institute for Visual & Analytic Computing, University of Rostock, Germany

Abstract

Time series segmentation is employed in various domains and continues to be a relevant topic of research. A segmentation pipeline is composed of different steps involving several parameterizable algorithms. Existing Visual Analytics approaches can help experts determine appropriate parameterizations and corresponding segmentation results for a given dataset. However, the results may also be afflicted with different types of uncertainties. Hence, experts face the additional challenge of understanding the reliability of multiple alternative the segmentation results. So far, the influence of uncertainties in the context of time series segmentation could not be investigated. We present an uncertainty-aware exploration approach for analyzing large sets of multivariate time series segmentations. The approach features an overview of uncertainties and time series segmentations, while detailed exploration is facilitated by (1) a lens-based focus+context technique and (2) uncertainty-based re-arrangement. The suitability of our uncertainty-aware design was evaluated in a quantitative user study, which resulted in interesting findings of general validity.

CCS Concepts

• ***Information systems*** → ***Uncertainty;*** • ***Human-centered computing*** → ***Visual analytics;*** *Visualization techniques;* • ***Computing methodologies*** → *Visual analytics;*

1. Introduction

Time series segmentation is the process of dividing time series into meaningful consecutive segments and assigning labels for the most likely activities represented by these segments. Before applying segmentation algorithms, the time series data requires multiple pre-processing steps—especially in the case of multivariate time series (MVTS). In order to tune these steps to a specific dataset, we employ multiple parameterizations and compute a large set of alternative time series segmentations—further called segmentation results. Röhlig et al. [RLK*15] use interactive analysis to evaluate segmentation pipelines using parameter-dependent exploration. The approach allows exploration of the label probabilities to identify areas where the segmentation algorithm showed low accuracy due to ambiguous activities being detected. The probabilities computed by the segmentation algorithm are one of several types of uncertainty introduced along pre-processing and modeling routines [BHJ*14, BBGM18], and more specifically segmentation pipelines (details on the segmentation pipeline used in this paper can be found in [BBB*18]). These approaches identified a need for uncertainty-aware exploration of the segmentation results. Multiple types of uncertainty need to be taken into account for an informed decision over the most appropriate segmentation result. Besides (1) uncertainties of the validity of segmentation results, i.e., label and segment probabilities, uncertainties may be (2) a-priori present in the data, or (3) might have been introduced by the segmentation pipeline (including data pre-processing).

One major obstacle to observing activities in a recorded MVTS is the high dimensionality. Dimensionality reduction can be applied [BHR*19], however, without contextualization this does not give any further insights into how to interpret the data. Alternatively, by combining an abstract time series representation with results from automated segmentation and labeling algorithms, the time series can be adequately partitioned for subsequent analysis. Moreover, capturing uncertainty introduced alongside the employed pipeline can be used as additional qualitative measures to be analyzed [BBB*19]. To this extent, the **main challenge** of **finding a good segmentation result a large number of results from the pipeline** is maintaining an overview while still considering different types of uncertainty is challenging. We break down these tasks into analyzing **(1)** the quality of a segmentation result for a specific parameter configuration, **(2)** the influence of uncertainty on segmentation results, specific time intervals, and individual timestamps, **(3)** the relation between types of uncertainty, and **(4)** the transition characteristics between segments. In their investigation, analysts more closely explore the computed segments, their individual lengths, the distribution of segments across the individual results, and the associated labels. Furthermore, the rate at which

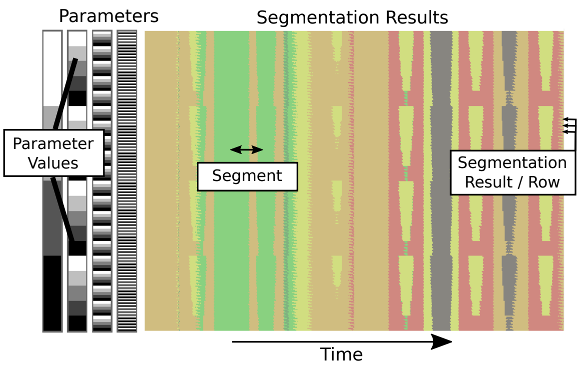
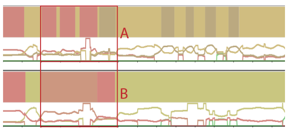
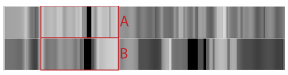


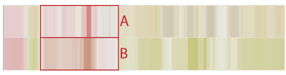
Figure 1: *Visualization of 300 segmentation results stacked on top of each other. On the right-hand side the computed segments are colored according to the assigned label. On the left-hand side the parameter values for 4 parameters are represented. The parameter grey scale ranges from dark for low values to light for high values.*



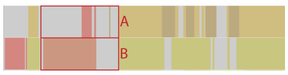
(a) *The regular segmentation result design: result uncertainty is shown in a line chart below the segments.*



(b) *The Uncertainty Heatmap design only encodes uncertainty as saturation.*



(c) *The Gradient Uncertainty design extends the regular segmentation result label view by encoding uncertainty as saturation.*



(d) *The Threshold Uncertainty design only shows segment labels for areas with uncertainty below a 40% threshold.*

Figure 2: *Visualization designs showing result uncertainty for uncertainty-aware segmentation result overview: the highlighted areas (A,B) show how time intervals were segmented by varying parameter settings and introduced uncertainty differently.*

label probabilities decrease and increase indicate algorithm sensitivity to detect segment transitions.

We propose a Visual Analytics (VA) approach that enables uncertainty-aware exploration of segmentation results (i.e., assessing the reliability of results) with respect to uncertainty a-priori present in the data and/or introduced by a segmentation pipeline. The contributions of this paper are as follows:

- Design and evaluation of uncertainty-aware visualizations to compare and assess large sets of alternative time series segmentation results (Section 2),
- Two lens-based Focus+Context (F+C) techniques for detailed uncertainty inspection (Section 3), and
- Uncertainty-based as well as lens-based re-arrangement of segmentation results (Section 3 & 2).

2. Visualization Design

The output of our segmentation pipeline consists of the segmentation and labeling results for each timestamp of the outgoing time series, the associated parameter settings that led to this result, and uncertainties introduced along the pipeline [BBGM18]. In this work we particularly focus on the visual encoding of two types of uncertainty: (a) *value uncertainty* is quantified either directly in the input data or it is externalized from the pre-processing operations employed in the segmentation pipeline (e.g., a smoothing operation to reduce noise). For every pre-processing step the introduced value uncertainty is recorded, resulting in a multivariate set of value uncertainties. (b) *Result uncertainty*, for example, can be represented as probabilities assigned by a labeling algorithm that gradually de- and increase over time, with the highest probability representing the winning label. Communicating these uncertainties enables the analyst to assess the reliability of the segmentation pipeline for a particular MVTS. Considering tasks **(1)**-**(4)** and the different types of uncertainties, we derived the following design rationales:

DR1 - Maintain a compact overview visualization: to show analysts a comprehensive overview of all results at once, computed with different parameter settings.
DR2 - Incorporate result uncertainty and value uncertainties in visualization design: to make analysts aware, so they can inspect if uncertainties had an influence on the results.
DR3 - Support different analysis goals with respect to uncertainty inspection: to let analysts explore, for instance, the probability of individual labels or the rate of transitions between labels.

The visualization is designed to show the entire set of segmentation results together with associated uncertainties in a compact representation (**DR1**). We extend the visualization design presented by [RLK*15, EST19]: The results of each parameter configuration are arranged vertically, with each line showing the different segments of one computed result over time. For a first overview of the results, labeled segments are displayed as colored bands without uncertainty visualization (see Figure 1). We augment this overview with *result uncertainty* for each timestamp. Inspired by the results of Gschwandtner et al. [GBFM16] we incorporate different ways of visualizing result uncertainty. Besides the visualization of segmentation results without uncertainty (see Figure 2a), the following designs can be toggled:

Uncertainty Heatmap Plot (Figure 2b)**:** To give an immediate impression of uncertainties only (no segment labels), we use a grayscale, with light values exhibiting high result uncertainty.
Gradient Uncertainty Plot (Figure 2c)**:** To show the different segments and labels together with their uncertainties, the default Gradient Uncertainty View uses color and transparency.
Threshold Uncertainty Plot (Figure 2d)**:** To enable a more directed exploration, this view shows only labels that exceed a certain probability threshold that is set interactively. This prevents visual clutter and facilitates the comparison of small differences between segmentation results.

We encode *value uncertainty* either as area (see Figure 3a) or brightness (see Figure 3b). The value uncertainty visualization can be inspected on demand (Figure 4 lens bottom) to explore the influence of value uncertainties on segmentation results.

(a) *Area Uncertainty plot encoding one dimension of value uncertainty.*

(b) *Uncertainty Heatband view encoding one dimension of value uncertainty.*

Figure 3: *Visualization designs for showing value uncertainty. When using (a) Area Uncertainty plots, value uncertainties from different sources can be stacked. An alternative visual encoding is the (b) Uncertainty Heatband plot. The two marked areas (A, B) show areas where value uncertainty is introduced periodically.*

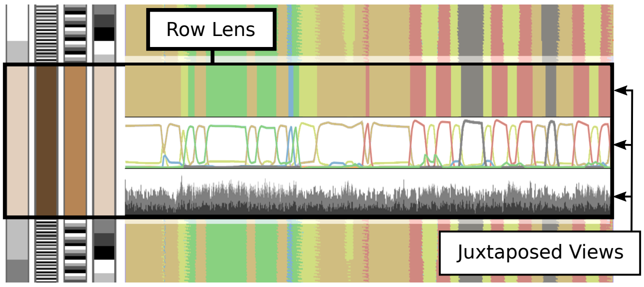


Figure 4: *The row lens allows for vertically zooming a segmentation result of interest. In this example, the lens area is used for showing three juxtaposed views displaying (from top to bottom) the segment label coloring , label probabilities as line plots, and the value uncertainty as stacked area plots. The parameter values are color-highlighted and provide detail information on mouse-over interaction.*

3. Interactions

To enable a detailed inspection of segment labels and uncertainties (**DR3**) we provide details on demand. Thus, we have designed two interactive lens views [TGK*17], the *row lens* and the *temporal lens*, to support exploration of both, result and value uncertainty on timestamp and segment level.

The **row lens** enlarges a single segmentation result to take up more vertical space. Depending on the zoom factor (defined by user interaction), we provide different composite visualizations (**DR2, DR3**). If vertical space does not permit multiple views, only one detail visualization is shown. Figure 4 (lens top) shows the distribution of segments and associated labels throughout the time series. Figure 4 (lens center) allows close inspection of the label distribution. Different representations can be toggled to show result uncertainties in detail. Figure 4 (lens bottom) plots value uncertainty over time so that analysts can relate value uncertainty influences to segments or intervals. The *row lens* does not distort the temporal axis (x-axis). This makes it easier to relate the currently magnified segmentation result to the remaining results. To also enable detailed inspection on a timestamp level, we designed the **temporal lens**, which magnifies both the vertical axis and the horizontal temporal axis. This allows us to display details about result and value uncertainties. Again, analysts can toggle between the different visual encodings. Figure 5 shows the lens focused on the transition between two different segments. Moving the cursor across segmentation results allows exploration of the underlying probabilities of labels and segments for each timestamp.

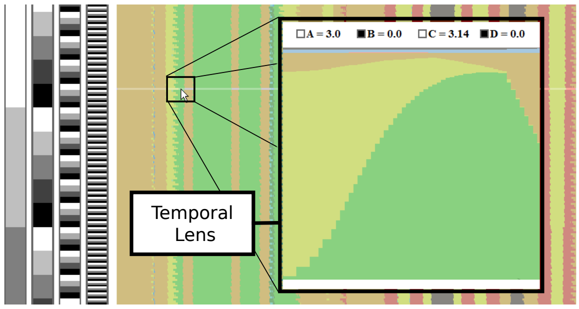


Figure 5: *A temporal lens that magnifies a segment boundary near the mouse cursor and shows the exact temporal change of label probabilities as a stacked area plot.*

Moreover, segmentation results might be mostly identical, with differences only surfacing in the associated result and value uncertainties, e.g., at the transitions between segments. In order to facilitate comparison of two segmentation results, it is sensible to position them in close proximity. We provide an **assisted re-arrangement** to address these issues. Segmentation results are **sorted according to their result uncertainties and value uncertainties**. The re-arrangement of rows is based on [EST19, RLK*15]: Rows in the overview visualization are permuted so that similar segmentation results are grouped to facilitate comparison. In our uncertainty-based re-arrangement approach, value and result uncertainties are used as sorting comparators. To determine the similarity of two segmentation results, uncertainty value differences are summarized for each timestamp. The smaller the sum, the more similar two segmentation results are to each other. In addition, we provide **lens-based re-arrangement**: The lenses can be used as a probing tool to identify and more effectively select interesting segmentation results and properties. If one or more interesting segmentation results were found through the lens, they can be marked as references for automatic rearrangement. The sorting algorithm then automatically reorders the rows based on the uncertainty types and temporal intervals, and segmentation result characteristics of the currently employed lenses.

4. Evaluation of Uncertainty Visualizations

We evaluated the appropriateness of the different visual encodings of uncertainty with a quantitative between-subject user study with 111 participants (details about the study and analysis of results can be found in the supplementary material). We provided participants with two or more alternative segmentation results for the same time interval and used different types of visual encodings of uncertainty, while participants had to choose the result with the highest or lowest uncertainty. We measured participants' error rate and task completion times to investigate the following hypotheses:

$\mathbf{H_0}$ The Gradient Uncertainty plot (Figure 2c) **does not perform worse** than the regular visualization of segmentation results as colored bars plus an additional line plot showing result uncertainty (Figure 4 lens top and center).

$\mathbf{H_1}$ The Gradient Uncertainty plot (Figure 2c) **does not perform worse** than the Uncertainty Heatmap plot (Figure 2b) showing result uncertainty.
$\mathbf{H_2}$ The Gradient Uncertainty plot (Figure 2c) **is more effective** than a Threshold Uncertainty plot (Figure 2d) for assessing result uncertainties of a large number of segmentation results, $\mathbf{H_{2a}}$ especially with limited vertical space available.
$\mathbf{H_3}$ The Uncertainty Heatband plot (Figure 3b) **does not perform worse** than the Area Uncertainty plot (Figure 3a) for showing value uncertainty.

From our study results we draw the following conclusions (considering the significance level of the p-value ≤ 0.05): (1) Hypotheses $\mathbf{H_0}$ and $\mathbf{H_1}$ could be confirmed which shows that the Gradient Uncertainty plot is not inferior to a composite view (showing the regular visualization of segmentation results as colored bars plus an additional line plot), although this designs provides more details. One reason for this might be that in the composite view segment label and uncertainty of this label had to be compared vertically instead of having both types of information encoded in one view. The Gradient Uncertainty plot is also not inferior to the Uncertainty Heatmap plot, although this design solely focuses on communicating uncertainty values and does not give additional information about labels.

(2) On the other hand, it could not be confirmed that Gradient Uncertainty plots outperform Threshold Uncertainty plots when showing a large number of segmentation results or when vertical space is limited ($\mathbf{H_2}$, $\mathbf{H_{2a}}$ rejected). The study results even show that for difficult scenarios with barely distinguishable uncertainties, the Threshold Uncertainty plot performed best (threshold mean error rate 47.75% vs. overall error rate 67.79%) This plot, however, comes also with the highest task completion times (median 26 seconds), probably as a consequence of the necessary interactions of adjusting the threshold. As a result, it is recommendable to employ different visualization designs that can be toggled, to support swift exploration using Gradient Uncertainty plots, and allow analysts to switch to Threshold Uncertainty plots for difficult cases.

Moreover, (3) the evaluation of value uncertainty designs showed equal scores and completion times for both the Heatband and Area Uncertainty plots ($\mathbf{H_3}$ confirmed). This extends designers' possibilities to more appropriately visualize this type of uncertainty. Favoring compactness of representation (**DR1**), we default to employing the Uncertainty Heatband in our *lens views*. This is particularly beneficial, if vertical space is limited: a heatband can be employed in the row lens with small height without any disadvantages for the user to assess value uncertainty.

5. Discussion

Analyzing large sets of time series segmentations is challenging, and conducting uncertainty-aware exploration further adds to the complexity of assessing the plethora of information provided. However, there are decisive benefits to uncertainty-aware analysis of results. Besides the obvious fact that analysts are provided with enough information to make informed decisions about the validity and credibility of segmentation results, there are several considerations we want to point out: (1) analysts have to assess whether the amount of uncertainty introduced into a segmentation result can be reduced. For example, a certain parameter makes segment transitions more accurate, reducing result uncertainty between labels. (2) The ability to estimate the influence of value uncertainties on the result outcome. By analyzing uncertainty introduced by preprocessing, analysts can determine how they affected the data and if the time series is still representative of the original data. If inappropriate parameter values are used in the segmentation pipeline, important patterns can be masked, rendering the segmentation ineffective, and (3) finding a good parameterization of the segmentation pipeline that can be applied to future time series data, the trade-off between an over-fitted model with zero uncertainty and just a bad segmentation with too much uncertainty has to be considered. Thus, some amount of uncertainty might even be desirable.

Our evaluation results showed increasing the complexity of the segmentation results view by adding uncertainty information did not decrease user performance. We found that the interactive Threshold Uncertainty plot showed the highest completion times (median overall: 19s, threshold: 26s), which we suspect was due to the necessary interactions users had to perform. Thus, to assess the influences of interactivity and limited vertical space on uncertainty-aware analysis, we aim to conduct a more specific user study on the appropriateness of Gradient Uncertainty and Threshold Uncertainty plots. Also the overall interactive design of our segmentation results visualization needs to be evaluated in future work. To further support users in the analysis of segmentation results, we are planning to integrate guidance methods [CGM*17]. This applies in particular to the use of lenses. Potentially interesting segmentation results could be automatically recognized (e.g., based on the uncertainty composition) and recommended for detailed inspection. This could be useful to unburden the analyst to some extent as well as to further simplify lens interaction.

6. Conclusion

We presented new visualization techniques to support uncertainty-aware exploration of large sets of time series segmentations. We visually communicate different types of uncertainties that stem from the data itself, the pre-processing of the data, and the uncertainty in the segmentation results, to foster awareness about the reliability of different results. We designed designated uncertainty visualizations as well as uncertainty-based sorting of results, and integrated these into the overview visualization of multiple alternative segmentation results to enable analysts to relate the segmentation result with different types of uncertainties that might have influenced it. We evaluated our visual encodings in a quantitative study, testing the performance of participants for a comparison task, which shed light on some interesting aspects that are relevant not only for the visualization of time series segmentation, but can be generalized to the visualization of similar uncertainty types with similar space constraints and tasks.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) and the Austrian Science Fund (FWF), Project No. I 2850 (-N31), Lead Agency Procedure (D-A-CH) "Visual Segmentation and Labeling of Multivariate Time Series (VISSECT)".

References

[BBB*18] Bernard J., Bors C., Bögl M., Eichner C., Gschwandtner T., Miksch S., Schumann H., Kohlhammer J.: Combining the Automated Segmentation and Visual Analysis of Multivariate Time Series. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2018), The Eurographics Association, pp. 49–53. URL: https://diglib.eg.org:443/xmlui/handle/10.2312/eurova20181112, doi:10.2312/eurova.20181112. 1

[BBB*19] Bors C., Bernard J., Bögl M., Gschwandtner T., Kohlhammer J., Miksch S.: Quantifying Uncertainty in Multivariate Time Series Pre-Processing. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2019), Landesberger T. v., Turkay C., (Eds.), The Eurographics Association, pp. 31–35. doi:10.2312/eurova.20191121. 1

[BBGM18] Bögl M., Bors C., Gschwandtner T., Miksch S.: Categorizing Uncertainties in the Process of Segmenting and Labeling Time Series Data. In *EuroVis 2018 - Posters* (2018), Puig A., Raidou R., (Eds.), The Eurographics Association, pp. 45–47. doi:10.2312/eurp.20181126. 1, 2

[BHJ*14] Bonneau G.-P., Hege H.-C., Johnson C. R., Oliveira M. M., Potter K., Rheingans P., Schultz T.: Overview and State-of-the-Art of Uncertainty Visualization. In *Scientific Visualization*, Mathematics and Visualization. Springer, London, 2014, pp. 3–27. URL: https://link.springer.com/chapter/10.1007/978-1-4471-6497-5_1, doi:10.1007/978-1-4471-6497-5_1. 1

[BHR*19] Bernard J., Hutter M., Reinemuth H., Pfeifer H., Bors C., Kohlhammer J.: Visual-Interactive Preprocessing of Multivariate Time Series Data. *Computer Graphics Forum 38*, 3 (2019), 401–412. doi:10.1111/cgf.13698. 1

[CGM*17] Ceneda D., Gschwandtner T., May T., Miksch S., Schulz H., Streit M., Tominski C.: Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (Jan. 2017), 111–120. doi:10.1109/TVCG.2016.2598468. 4

[EST19] Eichner C., Schumann H., Tominski C.: Making Parameter Dependencies of Time-Series Segmentation Visually Understandable. *Computers & Graphics* (2019). (accepted, to be published). 2, 3

[GBFM16] Gschwandtner T., Bögl M., Federico P., Miksch S.: Visual Encodings of Temporal Uncertainty: A Comparative User Study. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan. 2016), 539–548. doi:10.1109/TVCG.2015.2467752. 2

[RLK*15] Röhlig M., Luboschik M., Kruger F., Kirste T., Schumann H., Bögl M., Alsallakh B., Miksch S.: Supporting activity recognition by visual analytics. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Chicago, IL, USA, Oct. 2015), IEEE, pp. 41–48. URL: http://ieeexplore.ieee.org/document/7347629/, doi:10.1109/VAST.2015.7347629. 1, 2, 3

[TGK*17] Tominski C., Gladisch S., Kister U., Dachselt R., Schumann H.: Interactive Lenses for Visualization: An Extended Survey. *Comput. Graph. Forum 36*, 6 (Sept. 2017), 173–200. URL: https://doi.org/10.1111/cgf.12871, doi:10.1111/cgf.12871. 3