Eurographics Workshop on Visual Computing for Biology and Medicine (2019)
K. Lawonn and R. G. Raidou (Editors)

Short Paper

A Web-based Application for the Visual Exploration of Colon Morphology Data

J. Males[1], E. Monclús[1], J. Díaz[2] and P. Vázquez[1]

[1]ViRVIG Group, Universitat Politècnica de Catalunya, Spain
[2]Digital Care Research Group, Universitat de Vic - Universitat Central de Catalunya, Spain

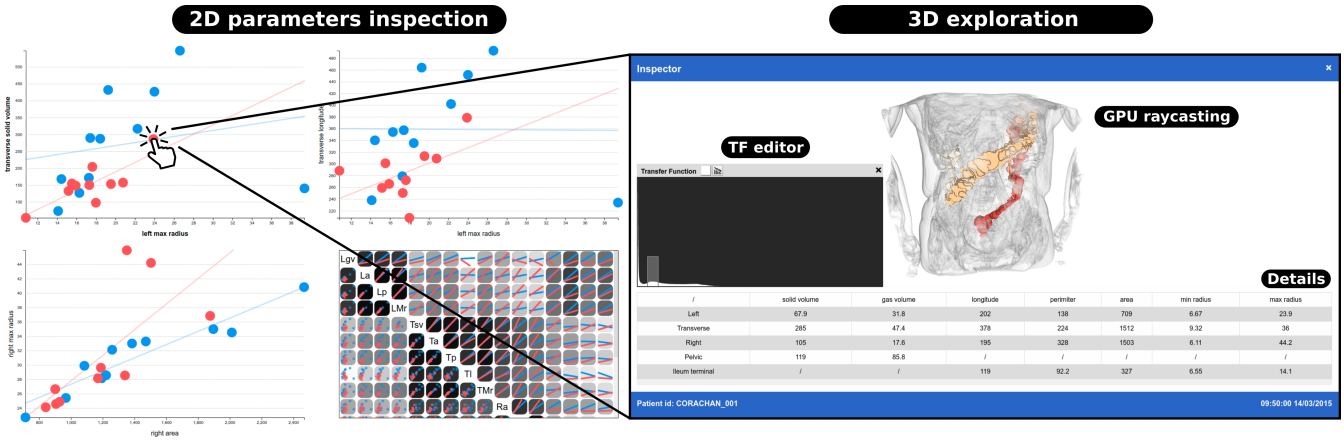

Figure 1: *Exploratory visual analysis of colon morphological data. The initial view (left) facilitates the exploration of the morphological measurements of the colon stored in the database. The two different colors (red and blue) indicate two different diet conditions at the moment of the data capture. The scatterplot matrix approach provides an overview and lets the user select the concrete plots to investigate. From the 2D scatterplots, the user can select a certain patient data, and the 3D exploration view (right) will automatically load the 3D model and the details (bottom). In this view, the user can inspect the model, as well as change the transfer function with the editor on the left.*

Abstract

The colon is an organ whose constant motility poses difficulties to its analysis. Although morphological data can be successfully extracted from Computational Tomography, its radiative nature makes it only indicated for patients with disorders. Only recently, acquisition techniques that rely on the use of Magnetic Resonance Imaging have matured enough to enable the generation of morphological colon data of healthy patients without preparation (i. e. administration of drugs or contrast agents). As a result, a database of colon morphological data for patients under different diets, has been created. Currently, the digestologists we collaborate with analyze the measured data of the gut by inspecting a set of spreadsheets. In this paper, we propose a system for the exploratory visual analysis of the whole database of morphological data at once. It provides features for the visual comparison of data correlations, the inspection of the morphological measures, as well 3D rendering of the colon segmented models. The system solely relies on the use of web technologies, which makes it portable even to mobile devices.

I.3.8 [Computer Graphics]: Applications— I.4.6 [Image Processing and Computer Vision]: Segmentation— J.3 [Life and Medical Science]: Health—

1. Introduction

The digestive system is very complex, with many organs and different functions that play an important role on our health. The colon is a continuously moving organ, with five different components (left, transverse, right, pelvic, and ileum terminal) that have largely varying morphologies among patients. Its size, length, volume, etc., does not seem to be related on other morphological parameters of the patients such as height or weight. Despite the continuous de-

DOI: 10.2312/vcbm.20191228

velopment of multiple imaging technologies, there is still a lack of knowledge on it, and one of the main reasons is the difficulty of gathering data. The imaging techniques commonly used for disorder diagnosis use contrast agents or drugs, a process known as patient preparation, and may include radiation methods, such as Computerized Tomography. Thus, understanding the behavior of the gut under different, normal conditions, such as a certain diet, has been highly challenging. The acquisition of colon data for non-prepared, healthy patients, with non-ionizing imaging techniques (i.e. Magnetic Resonance Imaging), has become a reality just very recently [BMM*17]. Thanks to this, we now have a database for the analysis of multiple patients' data, which may help establishing reference parameters to the *normal* gut behavior, and studying the effect of different diet conditions. The domain experts commonly explore these gathered data using spreadsheets, one per patient, which makes the process of extracting certain information quite cumbersome. More specifically, patient-to-patient comparison, and detailed inspection of the captured imaging is very costly with this kind of workflow, since it requires using several tools to fully explore the whole information of a single patient.

We propose a system that relies on the use of visual representations to show all the patients' data at once, with a customized SPLOM scatterplot matrix [HBO*10, ABSCW87] and individual scatterplots. Details on demand are accessed through a 3D view and the numerical data of the patients, as required by the user. This facilitates visual comparison as well as detailed inspection. All the development uses web technologies, making it available to any internet connected device that supports a modern browser.

2. Previous Work

The capture and analysis of multiple patients' data is an old problem, that can be addressed by statistical analysis or, more recently, with the use of visual tools (e.g. [AOH*14]). More concretely, the analysis of colon morphological data has been challenged by the difficulties in capturing the gastrointestinal organs with non-ionizing radiation [ASM15]. For the analysis of colonic contents a single MRI modality is not enough, since MRI-T2 images allow for the adequate segmentation of the colon boundary when there is presence of fat around the colon, but its contents cannot be extracted from it. MRI T1-weighted Fat-SAT (T1-FS) typically represents solid contents quite well, but gas and fat are poorly contrasted. A combination of such modalities can be used to successfully extract those [CMV*19], and build a database of morphological data.

Our approach deals with the analysis of multiple data at a time. Several visualization techniques have addressed this problem for morphological data under different approaches. Oeltze *et al.* [ODH*07] combine a set of 2D plots, 3D renditions, and small multiples to analyze perfusion data. Pastor *et al.* [PMT*15] also explore morphological data, but they concentrate into a multi-scale scheme for the 3D data. Preim *et al.* [PKH*16] present several techniques for multiple patient data inspection. In this case, the data also comes from medical image captures, but their system aims at extracting information from multiple parameters from a cohort of patients. Our approach, however, is more patient-centric, since what we want to explore is the data, and then go to the details of a single patient (in our case, the 3D view as well as the exact values of the measurements) on demand. It is therefore more similar to approaches such as the one by Keefe *et al.* [KERC09], where they inspect multiple data of biomechanical motion data and use small multiples to compare small motion sequences. Klimov *et al.* [KS05], like us, explore data of multiple patients under oncology treatment, but wiht an important emphasis on the time-based data exploration. Our 2D view design is similar to the approach taken by van den Elzen and van Wijk [vdEvW13], since we have all the time the small version of the scatterplots, and a subset of them in a larger version.

3. Visual Analysis of Morphology Data

The domain experts are used to analyze numerical data using spreadsheets, with one patient at a time. This makes the patient-to-patient comparison cumbersome, and relevant relationships or data cannot be inspected. The recently developed algorithms for colon data extraction in healthy patients (e.g. [CMV*19, OMB*18]) have opened the possibility of creating a database of real data whose morphology can be inspected to gain knowledge on the influence of diets [BMM*17] or prescription drugs [PMB*18], as well as other disorders. The domain experts we collaborate with have built a database of 21 patients with two different diets. To enable the exploratory visual analysis, we require a tool that easily facilitates the visual inspection of multiple parameters at once, and that lets the user go down to the captured 3D model, in case of need. Thus, our application has two major different components, a 2D chart-based tool, and a 3D volumetric exploration module. Another design goal decided early in the development was to make the application available in multiple platforms. Consequently, we used only web technologies for the implementation: HTML+JavaScript+D3 for the 2D data analysis, and JavaScript+WebGL for the 3D rendering.

2D layout. One of the main objectives of the project was to build a system that provides a quick and easy overview of the different morphological data from the patients. The purpose of such visualization is to facilitate the exploratory visual analysis, in search, for instance, of correlations between morphological measurements. Currently, little is known of the behavior of the different parts of the colon in healthy conditions in response to different diet conditions, for example, and no relationship has been established between different parameters of the colon, e.g. its length, with respect to other morphological characteristics of the patients, e.g. height or weight. Following the well known workflow of domain experts, we started with an initial representation for each pair of measures through a scatterplot, e.g. as the top left element in Figure 2. However, it was early noticed that this lets us see few comparisons, e.g. 4 at the same time. To rapidly analyze high amount of data, we decided to juxtapose all the relevant variables. So, we chose to render the morphological parameters in a SPLOM setup. Following the visualization mantra, we facilitate detailed inspection by providing full size charts, as shown in Figure 2, that can be selected from the SPLOM view by dragging the desired chart onto the corresponding view.

SPLOM design. The goal of the scatterplot matrix (SPLOM) component is to facilitate visual comparison, especially correlation finding, by juxtaposing several elements side-by-side. But a simple scaling and arranging of the data would waste a lot of space, as

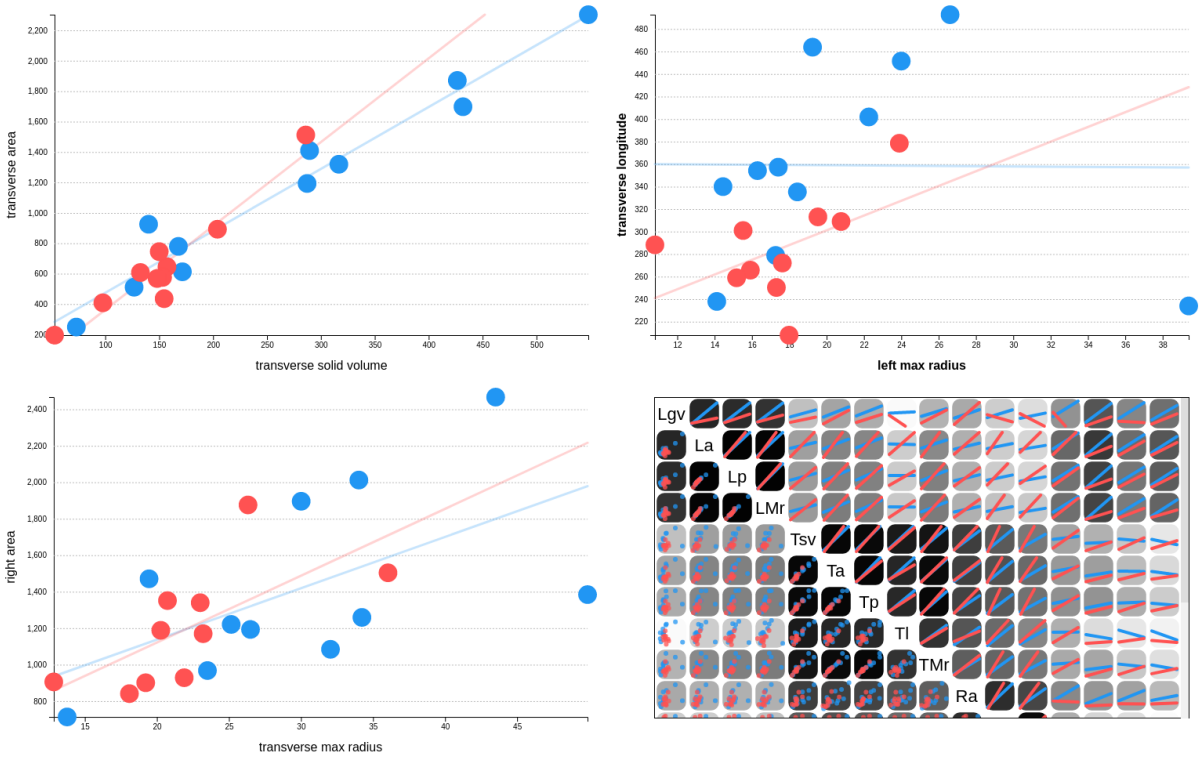

Figure 2: *2D layout. The application loads with the bottom right small plots view. By dragging any of the plots to the larger versions, the system loads the measures of the selected relationships. The SPLOM encodes the Pearson coefficient in the background (darker means higher correlation), and uses the top-right matrix part for the trend lines. The labels at the diagonal determine the parameters used in the plot.*

our matrix is symmetrical. Thus, we propose some improvements. First, instead of rendering the same plots twice, we separate scatterplot information from trend lines. Thus, we encode the scatterplots in the bottom-left part of the matrix, and the corresponding trend lines in the top right. Moreover, to facilitate the perception of the data, the layout of the chart is scaled down, but the plots are scaled up and rendered with transparency to make them visible and provide an accessible appearance of the points' distribution. For the top-right part, we designed a similar strategy: the whole chart is scaled down, but the line thickness is increased to facilitate legibility. One key element to study is the possibility of correlation between different morphological parameters. To facilitate the analysis, we calculate the Pearson correlation coefficient for each chart:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}} \quad (1)$$

where n is the sample size, x_i, y_i are sample values belonging to two sample datasets and $\bar{x}$, $\bar{y}$ are the samples' means. Then, we encode these values onto the charts themselves, by changing the background of the charts, from white (low or none correlation) to dark grey (high correlation). This makes reading the whole dataset easy, and the users can quickly go to the desired charts. To facilitate the identification of the parameters in the chart, displayed elements are encoded space efficiently with labels shown in the diagonal of the SPLOM. The labels *lab* consist on a set of letters that uniquely identify the words that describe the parameters with a fixed syntax: $< lab >:=< colon > [< modif >] < magnitude > [elem]$, where *modif* is a modifier that indicates the minimum or maximum value $(m|M)$ and *elem* determines whether solid or gas contents is the measured component $(g|s)$. The four different colons and the Ileum Terminal are encoded using the following symbols: **L:** Left, **T:** Transverse, **R:** Right, **P:** Pelvic, or **I:** Ileum terminal. Then, after the optional modifier, measured magnitudes are represented with: **r:** radius, **v:** volume, **a:** area, or **p:** perimeter. Finally, in some cases, when we refer to the content, we need to distinguish whether we are measuring gas or solid, so we use: **s:** solid, or **g:** gas. Thus, a label such as *Pv*, would refer to the pelvic volume, while *Tmr* refers to the minimum radius of the transverse colon.

Out of the 841 possible variable combinations, we only show the 120 most relevant in a 16×16 matrix. These are selected using the Pearson correlation matrix $A_{n\times n}$, where each element a_{ij} is the value of Pearson correlation coefficient between i and j colon variables, and a_{ii} is defined as 0. Over A we define a greedy algorithm that produces matrix $B_{m\times m}$, where $m \leq n$. We say that B contains the m-most significant variables of A. On every iteration of the greedy algorithm, we remove the least significant variable from A until only m columns and rows remain. The surviving entries in A are considered to be the most significant variables. The result is shown in Figure 3.

Interaction. The 2D view has some interaction features integrated

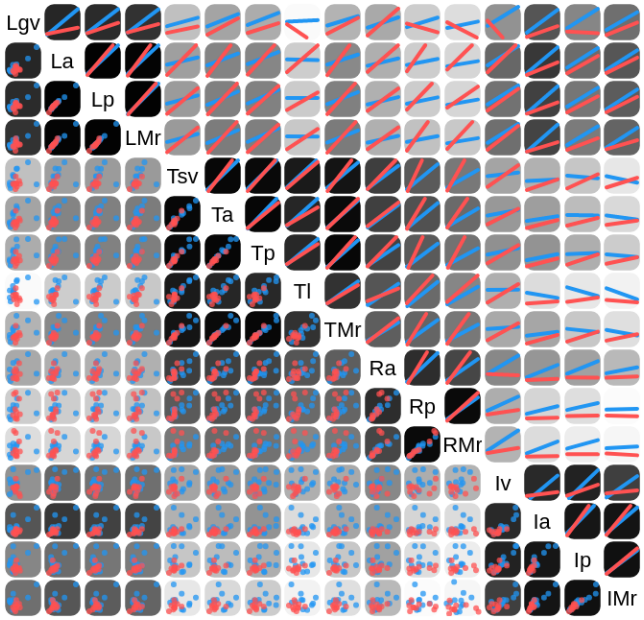

Figure 3: *SPLOM design. 16 (out of 41) different morphological measures are compared to each other. The bottom left triangle shows a simplified version of the scatterplots, while the top right shows the regression lines. The diagonal has been used to encode the labels to facilitate reading the values, and the background color encodes the Pearson coefficient which rapidly identifies the parameters with higher correlation. By dragging any of the plots over the larger charts, they are shown in detail.*

to facilitate the exploratory inspection. First, the user can drag the desired chart to the upscaled chart plots to facilitate the detailed inspection of the actual patients' data. Second, mouse hovering over the small plots triggers two visual cues: a blue label that verbosely indicates the parameters contained in the chart, and the corresponding plots, which are highlighted. This is carried out by adding an outline, on the chart the mouse hovered onto, as well as its symmetric counterpart, and changing the background to white. This facilitates the quick identification of both plots. By clicking on one of the small plots, it is transferred using an animated transition to one of the main, enlarged charts. The larger version of the plots also have some micro-animations to catch the attention of the user. First, when the mouse hovers a data point, it is enlarged. Second, for the line trends, its thickness is also increased on hovering.

3D Inspector Widget. The Inspector Widget (Figure 1 right) serves two purposes: access to details-on-demand, and 3D inspection of the patient model. For the details inspection, we provide a table with the captured morphology parameters of the different parts of the colon. The information of the patient (conveniently anonymized), the capture date and time are displayed at the bottom. The 3D model can be inspected through a typical ray-casting in the central part, that can be modified with a transfer function editor. As already mentioned, our ray-casting [KW03] is implemented using WebGL, which introduces some limitations regarding the computing capabilities at our disposal. To reduce the memory footprint, we use as input a floating point 3D texture containing the anatomical data and the segmentation information: positive values represent densities and negative ones for segmentation identifiers. Upon reading, density values are replaced by the corresponding color of each colon if the stored value indicates there is a selection. For the shading, we use a simple emission-absorption method with diffuse shading and a simple version of ambient occlusion based on the Starcraft's algorithm by Filion and McNaughton [FM08].

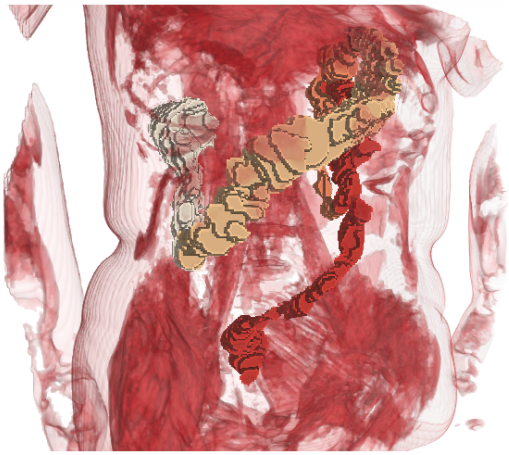

Figure 4: *Resulting image of the ray-casting method using the AO and diffuse lighting for the colons.*

4. Results

Performance was evaluated on an HP laptop ProBook 470 G5 with GNU/Debian buster/sid OS. The system is equipped with an Intel 64-bit quad-core i7-8550U processor, operating at 1.8 GHz with 8MB of cache. Furthermore, it has a GeForce 930MX NVIDIA graphics card with 2GB of memory and 16GB of working memory. We have tested our application in two different 64-bit browsers: Chromium 69.0.3497.92 and Firefox ESR 52.9.0.

When the website opens up, it generates not only the three plots, but every small plot in the bottom right cell. Even more, it sets up the entire inspector widget and its GPU ray caster behind the scenes. In a second phase, three main plots and the SPLOM load. To avoid lagging when the entering animations for the main plots and the matrix start, we only load main plots and display the right grid cell empty. Afterwards we spawn the matrix's elements. In a third phase, we construct the SPLOM. Firstly, the Pearson greedy algorithm for matrix dimensionality reduction is executed in order to cut down the number of plots that are going to be generated. Thereafter, the main grid cell gets populated by matrix elements and all of their entering animations play out to the end. All this setup process takes around 2.3 seconds. We also analyzed the memory consumption which does not grow higher than 24MB.

The WebGL rendering is more costly and depends on screen footprint, which may make framerates decay up to two times (e.g.

from 56 to 21 fps on our laptop testing machine) for zoomed views. In a desktop, framerates increase significantly.

Limitations. We adapted well-known visualization techniques, such as SPLOM, for our needs. Currently, we focus on finding correlations between the different measurements, although not all parameters are shown at once, which can be seen as a limitation. However, extending our data comparison for more variables is straightforward and could be implemented as a full-screen widget. Our SPLOM also requires a scroll bar because not all elements fit on the screen. Like in the previous case, this can be easily changed by resizing the components of the layout. The current views were designed to adapt to the positive aspect ratio of most screens. Another limitation comes from the model rendering. To make it interactive, we subsample the 3D texture, and thus, the quality could be improved. Faster WebGL implementations might allow us to increase the sampling rates [MKRE16, MF12], which would have a positive effect on the final images. The same can be said from the shading techniques: AO quality can also be refined by increasing the number of samples. Like in the previous case, we are limited by the performance of WebGL in the browsers we tested.

5. Conclusions and Future Work

In this work we presented our visual exploratory analysis tool for colon segmentation data sets. Our application is able to, not only summarize the given information, but also allow for dynamic examination of various variable combinations and grant detailed, on demand inspection of individual patient's measurements. Firstly, a matrix structure of SPLOM serves as an overview of the variable exploration space and gives hints on the data clusters that lie within. We used Pearson correlation coefficient to automatically reduce the matrix size and highlight correlative pairs of distinctive variables. The matrix also plays a role in the selection of the main charts' axial properties. Together with these scatter plots we allow the user to find linear correlations within the data set effortlessly. Furthermore, our application incorporates the Inspector Widget, which allows the user to visualize the segmented large intestine and shows patient measurements directly inside an overlook table.

As a future work, we would like to try out the presented application on different type of (medical) data sets. The project is highly extensible, since it essentially operates on raw data, other than the specific implementation of colon visualization. It would also be interesting to search for other kinds of correlation, e.g. by using clustering algorithms. Moreover, we are interested in improving the GPU ray caster with advanced shading and Transfer Function models. Finally, though the system was designed in collaboration with digestologists, no formal user study has been performed with domain experts. We would like to carry out some sessions with them to evaluate the utility and efficiency for them to analyze the data.

Acknowledgements

The authors want to thank the physicians of the Digestology Department at Hospital Vall d'Hebron. Work supported by the Spanish Ministry of Economy, Industry and Competitiveness and by the FEDER (EU) funds under the Grant No. TIN2017-88515-C2-1-R.

References

[ABSCW87] A. Becker R., S. Cleveland W., Wilks A.: Dynamic graphics for data analysis. *Statistical Science 2* (11 1987). 2

[AOH*14] Angelelli P., Oeltze S., Haász J., Turkay C., Hodneland E., Lundervold A., Lundervold A. J., Preim B., Hauser H.: Interactive visual analysis of heterogeneous cohort-study data. *IEEE computer graphics and applications 34*, 5 (2014), 70–82. 2

[ASM15] Alyami J., Spiller R., Marciani L.: Magnetic resonance imaging to evaluate gastrointestinal function. *Neurogastroenterology & Motility 27*, 12 (2015), 1687–1692. 2

[BMM*17] Bendezú R., Mego M., Monclús E., Merino X., Accarino A., Malagelada J., Navazo I., Azpiroz F.: Colonic content: effect of diet, meals, and defecation. *Neurogastroenterology & Motility 29*, 2 (2017), e12930. 2

[CMV*19] Ceballos V., Monclús E., Vázquez P., Bendezú A., Mego M., Merino X., Azpiroz F., Navazo I.: Colonic content assessment from mri imaging using a semi-automatic approach. In *VCBM 2019* (2019). 2

[FM08] Filion D., McNaughton R.: Effects & techniques. In *ACM SIGGRAPH 2008 Games* (2008), ACM, pp. 133–164. 4

[HBO*10] Heer J., Bostock M., Ogievetsky V., et al.: A tour through the visualization zoo. *Commun. Acm 53*, 6 (2010), 59–67. 2

[KERC09] Keefe D., Ewert M., Ribarsky W., Chang R.: Interactive coordinated multiple-view visualization of biomechanical motion data. *IEEE TVCG 15*, 11 (2009), 1383–90. 2

[KS05] Klimov D., Shahar Y.: A framework for intelligent visualization of multiple time-oriented medical records. In *Proc. AMIA* (2005), p. 405. 2

[KW03] Kruger J., Westermann R.: Acceleration techniques for gpu-based volume rendering. In *Proceedings of the 14th IEEE Visualization 2003 (VIS'03)* (Washington, DC, USA, 2003), VIS '03, IEEE Computer Society, pp. 38–. 4

[MF12] Mobeen M. M., Feng L.: High-performance volume rendering on the ubiquitous webgl platform. In *Proceedings of the 2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems* (Washington, DC, USA, 2012), HPCC '12, IEEE Computer Society, pp. 381–388. 5

[MKRE16] Mwalongo F., Krone M., Reina G., Ertl T.: State-of-the-art report in web-based visualization. *Comput. Graph. Forum 35*, 3 (June 2016), 553–575. 5

[ODH*07] Oeltze S., Doleisch H., Hauser H., Muigg P., Preim B.: Interactive visual analysis of perfusion data. *IEEE transactions on visualization and computer graphics 13*, 6 (2007), 1392–1399. 2

[OMB*18] Orellana B., Monclús E., Brunet P., Navazo I., Bendezú Á., Azpiroz F.: Quasi-automatic colon segmentation on t2-mri images with low user effort. In *MICCAI* (2018), Springer, pp. 638–647. 2

[PKH*16] Preim B., Klemm P., Hauser H., Hegenscheid K., Oeltze S., Toennies K., Völzke H.: Visual analytics of image-centric cohort studies in epidemiology. In *Visualization in medicine and life sciences III*. Springer, 2016, pp. 221–248. 2

[PMB*18] Poulsen J. L., Mark E. B., Brock C., Frøkjær J. B., Krogh K., Drewes A. M.: Colorectal transit and volume during treatment with prolonged-release oxycodone/naloxone versus oxycodone plus macrogol 3350. *Journal of neurogastroenterology and motility 24*, 1 (2018), 119. 2

[PMT*15] Pastor L., Mata S., Toharia P., Bayona S., Brito J. P., Garcia-Cantero J. J.: Neuroscheme: efficient multiscale representations for the visual exploration of morphological data in the human brain neocortex. 2

[vdEvW13] van den Elzen S., van Wijk J. J.: Small multiples, large singles: A new approach for visual data exploration. *Computer Graphics Forum 32*, 3pt2 (2013), 191–200. 2