Interactive framework for the visual exploration of colonic data

Jan Males, Eva Monclús, Jose Díaz, Isabel Navazo, Pere-Pau Vázquez

VIRVIG Group, Universitat Politècnica de Catalunya, Barcelona, Spain
Digital Care Research Group, Universitat de Vic - Universitat Central de Catalunya, Vic, Spain

ARTICLE INFO

Article history:
Received June 25, 2020

Keywords: Visualization, Web-based application, Volumetric data

ABSTRACT

Computerized Tomography (CT) and, more recently, Magnetic Resonance Imaging (MRI) have become the state-of-the-art techniques for morpho-volumetric analysis of abdominal cavities. Due to its constant motility, the colon is an organ difficult to analyze. Unfortunately, CT’s radiative nature makes it only indicated for patients with important disorders. Lately, acquisition techniques that rely on the use of MRI have matured enough to enable the analysis of colon data. This allows gathering data of patients without preparation (i.e. administration of drugs or contrast agents), and incorporating data of patients with non life-threatening diseases and healthy subjects to databases. In this paper we present an end-to-end framework that comprises all the steps to extract colon content and morphology data coupled with a web-based visualization tool that facilitates the visual exploration of such data. We also introduce the set of tools for the extraction of morphological data, and a detailed description of a specifically-designed interactive tool that facilitates a visual comparison of numerical variables within a set of patients, as well as a detailed inspection of an individual. Our prototype was evaluated by domain experts, which showed that our visual approach may reduce the costly process of colon data analysis. As a result, physicians have been able to get new insights on the effects of diets, and also to obtain a better understanding on the motility of the colon.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The digestive system is very complex, with many organs and different functions that play an important role on our health. Among all the organs inside the abdominal cavity, we are interested in the colon. The colon is a continuously moving organ, with four different segments: left (or ascending), transverse, right (or descending), and pelvic. These segments have largely varying morphologies among patients: their size, length, volume, etc., do not seem to be related to other morphological parameters of the patients such as height or weight. Despite the continuous development of multiple imaging technologies, there are some aspects, such as the effect of diets on the colon content or its motility, that have not yet been studied. 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 (CT). This poses some difficulties: first, using CT on healthy subjects or patients with no life-threatening pathologies is discouraged, and it has been eventually limited by Ethics Committees in hospitals, due to the high number of scans required. Second, patient preparation also results in new variables to consider, and thus, it is another source of noise that may collide with the diagnosis. The acquisition of colon data for non-
prepared patients, with non-ionizing imaging techniques (i.e.
Magnetic Resonance Imaging, MRI [11]), has become a reality
just very recently. The use of MRI opens the door to the cre-
ation of databases that include healthy subjects. This facilitates
the analysis of multiple patients’ data and helps establishing ref-
erence parameters to the normal gut behavior, and also permits
studying the effect of different diet conditions, or the influence
of diseases in the motility of the colon.

Our research was guided by our collaborators’ need to im-
prove their work in two directions: first, being able to gather
data in a more efficient way. Second, improving their data
analysis process that now takes several hours with the common
tools. This data analysis must allow them to get a better un-
derstanding of two main factors: i) the relationship between
different morphological parameters of the colon in order to de-
terminate whether there is any correlation within them in a group
of patients, and ii) analyzing the presence of intestinal motil-
ity under certain affections, such as cystic fibrosis. Obviously,
other patients without life-threatening diseases can also be ana-
lyzed with this approach, and improve their life conditions.

This work is an extension of our previous paper in the Eu-
rographics Workshop on Visual Computing for Biology and
Medicine [2], where we presented a first prototype of the vi-
sual analysis tool (module $M_{visPro2D}$ shown in Figure 1). Our
contributions can be summarized as follows:

- A set of tools for the extraction of morphological measure-
ments from MRI images (Section 3.2).
- A framework for efficient data extraction and visual anal-
ysis of morphological and contents data from the colon
using MRI images (Section 3.1).
- A web-based tool for the visual exploration of morphologi-
cal data (Section 4), and an extension to this initial proto-
type to include motility data obtained through a different
 gathering protocol (Section 7.1).
- A details-on-demand widget, that includes a WebGL ren-
dering of the segmented volume (Section 5).
- An evaluation of our system by domain experts (Sec-
ction 8).

Our first contribution consists in the creation of a software
that facilitates the extraction of morphological data with little
invasive imaging techniques. Previous research allowed us to
extract colonic content data, but lacked the information on the
morphology of the colon. Our second contribution is a full
framework composed of a set of modules that facilitates the
whole data extraction and its visual analysis. By reusing the
same input necessary to perform the data segmentation, physi-
cians can now extract both content and morphological data.
With this new system, databases of patients can now be created
with much less effort.

Typically, domain experts gather data from two different pa-
tient groups, with two different conditions. Then, the data is
analyzed using statistical software (Excel and SPSS), and go-
ing down to the real data (e.g. exploring the geometry) for cer-
tain cases. Unfortunately, to fully explore the data and extract

conclusions, physicians require from three to six hours, mostly
dependent on the number of variables involved. Thus, our third
contribution lies in the creation of an interactive visual tool to
facilitate the exploratory analysis of the extracted data through a
customized scatterplot matrix design [3,4] and individual scat-
terplots.

Our approach may reduce this costly process by providing a
specifically-designed tool that eases the analysis of multiple
variables, thus boosting the initial data exploration. Essentially,
the domain experts want to be able to quickly assess whether
certain variables present correlations within a set of patients. If
that is the case, they would like to see which variables, and in-
spect the patient data, for instance, by getting a 3D view of the
segmented colon together with the information of the numeric
extracted data. To do this, we provide details-on-demand in-
spection in the form of tables or a simple MRI viewer in WebGL
within the same tool. This facilitates visual comparison as well
as detailed inspection. The visualization tool uses web tech-
nologies, so the exploratory data visualization can be carried
out with any internet connected device that supports a modern
browser. We developed a web-based tool, instead of a desktop-
based application because the installation of third-party soft-
ware is strictly limited in the shared PCs of the hospitals. Previ-
ous collaborations with the same group of professionals ended
in customized installations in the physicians’ personal laptops,
thus limiting the possibilities of collaboration with their col-
leagues.

Note that we use two types of data from the colon: morpho-
logical (or geometric) data and colonic content data. Both are
extracted through different procedures, as we explain in the pa-
er. Our initial data visualization tool was evaluated informally
by three domain experts in a demo session, where they sug-
gested some small modifications, and an extension to facilitate
comparison with other external data obtained through wireless
video capsule endoscopy [5]. As a result, we had a third data
source that we call motility data. After implementing a second
version of the prototype for the visual comparison of colon con-
tent with motility data, the new web application was evaluated
formally by two domain experts.

The rest of the paper is organized as follows: the next sec-
tion reviews the previous work. In Section 3 we introduce our
data extraction pipeline and describe the different algorithms
for obtaining the colonic data later used for analysis, as well as
the database construction protocol. Section 4 presents the web-
based tool for exploratory data analysis and Section 5 describes
the 3D rendering and details-on-demand module. In Section 6
we describe the initial user study. As a result of this study, we
modified and extended our visual exploration tool to deal with
motility data, described in Section 7. Section 8 presents the
evaluation of the second prototype. Finally, Section 9 discusses
the results and concludes the paper pointing out some lines of
future research.

2. Previous Work

This section briefly reviews the work that is most closely re-
lated to our approach from the medical point of view. For ad-
comes from medical image captures, but their system aims to for multiple patient data inspection. In this case, the data also preim et al. for the 3D data. Preim... several visualization techniques have addressed this problem with the exploratory analysis of multiple variables at a time. Our approach deals with the exploratory analysis of multiple variables at a time. Several visualization techniques have addressed this problem for morphological data under different approaches. For example, Oeltze et al. [15] combine a set of 2D plots, 3D renditions, and small multiples to analyze perfusion data. Our approach has similar components, but our workflow is completely different, since our main goal is the visual comparison of the extracted data from multiple patients. Pastor et al. [16] also explore morphological data, but they concentrate on a multi-scale scheme for the 3D data. Preim et al. [17] present several techniques for multiple patient data inspection. In this case, the data also comes from medical image captures, but their system aims to extract information from multiple parameters from a cohort of patients. Therefore, a great effort has to be put into clustering the dataset to extract/form hypotheses, while we can show and inspect patients individually. Our approach, however, is more patient-centric, since what we want to explore is the data, and then go to the details of a patient (in our case, the 3D view and the exact values of the measurements) on demand. It is therefore more similar to approaches such as the one by Keefe et al. [18], where they inspect multiple data of biomechanical motion data and use small multiples to compare small motion sequences. Klimov et al. [19], like us, explore data of multiple patients. In their case, since the patients are under oncology treatment, an important emphasis is put onto the time-based data exploration. Our 2D view design is similar to the approach taken by van den Elzen and van Wijk [20], since we always have the small version of the scatterplots, and a subset of them in a larger version. Yates et al. [21] address the problem of very large SPLOMs by designing a visualization that attempts to summarize the SPLOM contents. Finally, Matute et al. [22] also reduce the required space by simplifying the scatterplot contents using a skeleton to represent the spatial distribution.

3. Colonic content and morphology analysis

As outlined above, we have created a framework for the extraction and visual analysis of the colon content and morphology data. In this section we present the framework and our new geometry extraction module. In Section 3.3, we explain how it has been successfully used to build databases of patients and healthy subjects to gather new knowledge on the effects of diets in the colon.

3.1. Framework architecture

Our framework consists of a set of modules that provide two main features: data extraction and visual exploration. The overall architecture can be seen in Figure [1]. From the T2 images, we...
use a quasi-automatic segmentation technique (module \( M_{\text{SegT2}} \)) proposed by Orellana et al.\(^\text{(23)}\) to segment the colon voxels.

The output (\( \text{colonSegT2} \)) is used along with the T1-FS images to extract the colon content information (gas, solid, and semi-solid), by applying the approach presented by Ceballos et al.\(^\text{(13)}\): first, we register the segmented T2 volume to the non-segmented T1-FS, and then, we proceed to segment T1-FS images. This happens in module \( M_{\text{SegT1-FS}} \). In a second step (depicted as \( M_{\text{Content}} \)), this information (\( \text{colonSegT1-FS} \)) is used to extract the data of the contents at the different segments of the colon. To fully understand the changes that happen to the colon under different diets or diseases, besides colon content, medical experts suggested to incorporate information on the colon morphology. This kind of information had not been analyzed before. Hence, we have designed and implemented a new module (\( M_{\text{Geom}} \)) that takes as input the results of the T2 segmentation data and extracts geometric features, as explained in the following section. The remaining modules are the ones for data visualization: first, we have the 2D exploratory visualization module (\( M_{\text{VisPlot2D}} \)), which lets users search for correlations among many patients’ data. Second, we have the \( M_{\text{VisPlot3D}} \) module to access details-on-demand of the individual patients data. This is done by providing an overview-and-detail view that shows the 3D volumetric data of the colon, as well as the extracted numeric data under analysis.

### 3.2. Extraction of Morphological data from MRI

Gastroenterological disorders and digestive dysfunctions affect the content, motility, and, it is hypothesized, also the morphological shape of the colon. Therefore, their measurement, evaluation, and correlation may have an important influence in the diagnosis.

Our approach to compute the colon morphological properties relies upon an accepted practice in medicine, which assumes that the colon anatomy can be modeled as a generalized tube with a variable radius along each segment of the colon and the ileum. The parameters that physicians would like to analyze for the left, transverse, right, and pelvic segments of the colon, as well as for the ileum are:

- The length.
- The minimum, maximum, and average radius.
- The area and the perimeter (calculated as described below).
- The volume.

In order to obtain this information, we have developed a new module, named \( M_{\text{Geom}} \). It gets as input data the information provided by the module \( M_{\text{SegT2}} \): the colon segmentation from T2 images (\( \text{colonSegT2} \)), and the user input (marker points).

The module for colon segmentation from unprepared MRI (\( M_{\text{SegT2}} \)) requires that the specialist provides a minimal set of five anatomical reference points (marker points) along the colon trajectory to guide the segmentation algorithm, that is why we call it quasi-automatic. These points determine the start and end of each of the colon segments: left, transverse, right, and pelvic. Depending on the complexity of the colon shape, more points may be required to ensure a correct segmentation result.

Our morphology extraction algorithm will reuse this input, and performs its calculations automatically. The final colon segmentation is a 3D discrete model where each voxel has an identifier/label corresponding to the colon segment it belongs to.

Our process to automatically and efficiently obtain the morphological properties, illustrated as \( M_{\text{Geom}} \) in Figure\(^\text{1} \) consists of two steps: i) medial path extraction, and ii) geometric properties extraction. The result is a set of variables that, according to the experts, are enough to capture the geometrical variation of the colon between the different patients involved in the study. These methods work as follows:

**Medial path extraction.** First, we compute the medial path (a set of 3D polylines) of the colon. We start with a 3D discrete skeletonization of the segmented voxel model. As it is well known in the literature\(^\text{(24)}\), the obtained skeleton can have an irregular shape due to slight irregularities in the object boundary, which may interfere with processes based on the topological properties of the skeleton. That can be avoided by reusing the list of sorted points (marker points) provided by the specialist along the colon in the module \( M_{\text{SegT2}} \). These are used to compute the shortest path from two consecutive points in the marker points using the obtained skeleton. This step is performed by building a graph from the skeleton points and running Dijkstra’s shortest path algorithm on it. In this way, a one-dimensional set of connected voxels is obtained, and a set of 3D polylines (one for each colon segment) is extracted from it.

**Geometric properties calculation.** Once we have computed the medial colon path, we can proceed to measure all the geometric properties related to the generalized tubular structure model of the colon. The length of each segment is the length of the polyline that models the axis of its tubular model (Figure\(^\text{2} \)(b)). The volume is estimated by adding the volume of the voxels of each segment. Since the radius of the tube is variable along each segment, we proceed to calculate cross-sections along different sampling positions of the 3D polylines, as illustrated in Figure\(^\text{2} \)(c). This lets us obtain the minimum, maximum, and the average values for the tube radius, and the area and perimeter, using standard image processing methods.

Area and perimeter are defined as the average of these values obtained at the computed cross-sections at those sampling positions. Note that, when at a given point of the polyline the cross-section of the tube is computed, it may intersect different sections of the colon segmentation. The correct one is the one that contains the selected point.

The main novelty of our technique is that, to our knowledge, it is the first one able to extract colon morphology information from unprepared patients. Moreover, we also know that the accuracy of measurements is good enough for the clinical practice, because we obtain the segmented images with a previously validated process\(^\text{(25)}\).

For the visual analysis we also incorporate colon content data obtained from module \( M_{\text{SegT1-FS}} \). This module automatically calculates the faecal colonic content from T1-FS images, by means of a classification step\(^\text{(13)}\) based on a k-means clustering approach that classifies all the voxels inside the colon.
3.3. Acquisition protocol

With our system.

3.3. Database construction

Segmentation as: non-solid, semi-solid or solid.

3.3.1. Acquisition protocol

The parameters used in the T2 acquisition for the group A were the following ones: matrix resolution = 256 × 256. In-plane resolution = 1.7 × 1.7 mm. Slice thickness = 3.5. Number of slices = 50. For the group B, the parameters were: matrix resolution = 256 × 244. In-plane resolution = 1.67 × 1.67 mm. Slice thickness = 3.5. Number of slices = [40, 54].

In the T1-FS acquisition process, the parameters used for the group A were: matrix resolution = 270 × 320. In-plane resolution = 1.5 × 1.5 mm. Slice thickness = 1.5 mm. Number of slices = 120. Instead, the parameters for the group B acquisition were: matrix resolution = 320 × 210. In-plane resolution = 1.75 × 1.75 mm. Slice thickness = 1.5 mm. Number of slices = [40, 128].

It must be remarked that no oral or intravenous contrast neither antiperistaltic drugs could be used in any of the imaging protocols, as determined by the clinical experts. All images were codified and analyzed blinded concerning the source, acquisition, and any preceding intervention by two trained physicians in MRI images under the supervision of an expert radiologist.

3.3.2. Data extraction procedure

In this section we describe how physicians gather the data from MRI scans, acquired in the scope of different clinical studies (group A and B). In all the experiments, each time a subject was scanned, two coronal image series of the abdomen were obtained: one T2-weighted HASTE sequence during two apneas of approximately 20 seconds each, and a T1-weighted VIBE Fat-Sat sequence in one apnea of 12 seconds.

The imaging examinations were performed on 1.5-T MR imaging systems with two six-channel phased-array abdominal coils to cover the whole abdomen. The subjects were placed in prone position with a 4-element body coil wrapped around the abdomen. The acquisition protocols for the two groups were different.

From a clinical perspective, the group A was collected in a study about the effects of the diet on the colonic content [25]. It consisted of 30 MRI scans acquired from 15 healthy adult (>18 years) volunteers after and before the defecation. Moreover, a subgroup of 10 participants were instructed to follow a 3-day low-residue diet and a 3-day high-residue diet with a 3-day diet-free interval. In the morning immediately after each period, a...
couple of abdominal MRI scans (T2 & T1-FS) were acquired after an overnight fast to measure colonic content. The subjects in this study also participated, as it is shown in Section [7], in a second clinical study focused on the relationship between motility and cystic fibrosis (see Section [7,2]).

This database of healthy subjects has allowed our collaborators to analyze the colon behavior. This led them to obtain several new findings. For example, it was determined that the total turnover of the colon is very quick (around few hours): over one-third of colonic content was cleared daily and replaced, indicating that this is a very dynamic process. They also found that the fractional clearance rate of colonic content was not influenced by the diet or colonic content volume. Comparison of colonic content before and after defecation also indicates that the process of defecation involves not only rectal emptying but also a major redistribution of contents along the colon. Other findings related to the variation of the contents and the amount of gas according to the diet were also found [25].

We performed a first evaluation of the exploratory visualization tool with the aid of medical experts using a subgroup of 13 subjects of group A (healthy subjects) and all subjects in group B (cystic fibrosis patients). The complete database of group B was used in our extended prototype (see Section [7]).

4. Visual Analysis of Morphology Data

As already explained, the domain experts we collaborate with are used to analyze numerical data using spreadsheets with Excel and SPSS. The process takes them up to three hours for the database of morphology data. Our goal is to develop a tool to facilitate this task, as well as the detailed analysis of individual cases or patient-to-patient comparison without the need of using extra software. Therefore, our initial approach is to build a tool to analyze the recently built database of real data with the goal of gaining knowledge on the influence of diets [25]. Our intention is to group in a single web-based tool, the statistical analysis features they use in SPSS and the possibility of individual inspection of numerical and 3D data for patients.

To enable the exploratory visual analysis, we have created a visual inspection tool with the following requirements: i) users must be able to analyze multiple parameters at once, ii) users should be able to go to the real data, that is, the captured 3D model, in case of need, and iii) the application must be easily accessed from different computers.

As a result, we designed an application with two coupled components: a 2D chart-based tool for numerical data analysis, and a 3D volumetric exploration module that serves for individual patient analysis. To fulfill the third design goal, we chose to use only web technologies for the implementation: HTML+JavaScript+D3 for the 2D data analysis, and JavaScript+WebGL for the 3D rendering. This overcomes the limitations of third-party software installation present in hospitals for desktop PCs.

4.1. 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 extracted from the acquired sets. The purpose of such visualization is to facilitate the exploratory visual analysis, in search, for instance, of correlations between morphological measures. Currently, little is known about the behavior of the different parts of a healthy colon in response to different diet conditions. 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 [3]. However, it was early noticed that this lets us see few comparisons, roughly four at the same time on a laptop screen. To rapidly analyze higher amounts of data, we decided to juxtapose all the relevant variables. So, we chose to render the morphological parameters in a scatterplot matrix (SPLOM) setup. Following the visualization mantra, we facilitate a detailed inspection by providing full size charts, as shown in Figure [3] that can be easily selected from the SPLOM view by simply dragging the desired chart onto the corresponding view. The 2D scatterplot is standard, but the SPLOM has been modified with respect to what we find in the literature to better address our domain.

4.2. SPLOM design

The goal of the scatterplot matrix component is to facilitate visual comparison, especially correlation finding, by juxtaposing several elements side-by-side. A simple scaling and arranging of the data would waste a lot of space, since our matrix is symmetrical. Therefore, we propose some improvements over this initial design. First, instead of rendering the same plots twice, we differentiate between scatterplot information and trend lines. Thus, the scatterplots (displayed on the bottom-left part of the matrix) appear separated from the regression lines (on the top right). Each color indicates a patient group. Moreover, to facilitate the perception of the data, the size of the chart is scaled down, but the plots are scaled up and rendered with transparency to make them visible and provide a clear enough 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 a correlation between different morphological parameters. To facilitate the analysis, we calculate the Pearson correlation coefficient for each chart:

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

(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. This extra information has been added by changing the background of the charts, from white (low or no correlation) to dark grey (high correlation). This makes reading the whole dataset easy, and the users can quickly go to the desired charts.
Fig. 3. Overview of the 2D representation. The layout shows the SPLOM on the bottom right. The user can then see the different relationships between the measures taken from the different colon segments. By dragging any of the plots to the larger versions, the system loads the corresponding data. The SPLOM encodes the Pearson coefficient in the background (darker color higher correlation), and codifies trend lines in the top-right matrix part of the SPLOM view to facilitate the interpretation of the data. Trend lines are computed with the least squares approach, but other, more sophisticated techniques could be used. Our collaborators normally use only this simple version, but adding other approximations would be straightforward. The labels at the diagonal determine the parameters used in the plot, which saves space. The color of dots and lines (blue and red) indicate the patients’ groups (A and B, respectively).

In order to facilitate the identification of the parameters represented in a chart, we encode in a space-efficient way the elements displayed. We do so by placing the required labels in the diagonal of the SPLOM. This saves space at the top and left of the charts and facilitates reading. The labels consist of a set of letters. These uniquely identify the words that describe the parameters with a fixed syntax:

\[ \text{< label >:< colon > [< modif >] < magnitude > [elem]}, \]

where \text{modif} is a modifier that indicates the minimum or maximum value (\text{mM}) and \text{elem} determines whether solid or gas contents is the measured component (\text{g/s}). The colon segments and the Ileum Terminal are encoded as: \text{L:} Left, \text{T:} Transverse, \text{R:} Right, \text{P:} Pelvic, or \text{I:} Ileum terminal. Then, after the optional modifier, measured magnitudes are represented with: \text{r:} radius, \text{v:} volume, \text{a:} area, or \text{p:} perimeter. Finally, when we refer to the content, we distinguish whether we are measuring gas (\text{g}) or solid (\text{s}). In this way, a label such as \text{Prv}, would refer to the pelvic volume, while \text{Tmr} refers to the minimum radius of the transverse colon.

Out of the 841 possible variable combinations, we only show 120 using the bottom left triangle of a $16 \times 16$ matrix (the upper right triangle shows the trend lines, and the diagonal the labels). Each plot represents one of the variable pairs with higher maximum Pearson correlation, as explained next, since this is what physicians were interested in regarding this dataset. In order to determine the variables to represent, we use the Pearson correlation matrix $A_{n,n}$, where each element $a_{ij}$ is the value of the Pearson correlation coefficient between the $i$ and $j$ variables, for $i \neq j$. Using $A$ we define a greedy algorithm that produces matrix $B_{m,m}$, where $m \leq n$. Let $B$ be the set of the $m$-most significant variables of $A$. On every iteration of the greedy algorithm, we remove the variable with lower maximum correlation 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 4.

The resulting application, at full screen, is clear enough to convey the point distributions in the SPLOM for the current sample sizes. If the number of patients was in the order of hundreds, the charts would probably become cluttered. Fortunately, with the type of data we are dealing with, this is an unlikely event. This is due to two main reasons: first, the number of patients a physician sees with a certain condition (e.g. cystic fibrosis) is limited. Second, the data extraction procedure is
values, and the background color encodes the Pearson coefficient. This facilitates the rapid identification of the variables set with a higher correlation coefficient. The color of dots and lines (blue and red) indicate the patients’ groups (A and B, respectively). By dragging any of the plots over the larger charts, they are shown in detail.

4.3. Interaction

The 2D view has some interaction features integrated to facilitate the exploratory inspection. First, as already commented, the user can drag the desired chart to the larger chart plots to facilitate the detailed inspection of the actual patients’ data. Second, mouse hovering over the small plots triggers two visual cues: a label that verbosely indicates the parameters plotted in the chart, and the corresponding plots are highlighted. This is carried out by adding an outline, on the chart the mouse hovered onto, as well as its symmetric counterpart. This facilitates the quick identification of both plots, as shown in Figure 5. By clicking on one of the small plots, it is transferred using an animated transition to one of the main, enlarged charts.

The larger versions 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.

Detailed inspection of all the data of a patient can be achieved through the Inspector Widget, as explained in the following section.

5. Inspector Widget

The Inspector Widget (see Figure 6) serves two purposes: first, access to details on-demand such us the longitude, perimeter or area of each colon segment. Second, visualize and allow the 3D exploration in real-time of the patient’s data. Because of the large number of captured parameters (41 values), morphology data are presented in a table (Figure 6a), while the name of the patient (conveniently anonymized in Figure 6b), the capture date and time are shown at the bottom (c). The central part of the widget (d) displays the rendering of the patient’s colon that can be explored and modified by means of the simple transfer function editor located on the left (b). The latter allows to create 1D transfer functions for applying colors and opacities to data samples. Direct Volume Rendering (DVR) using the well-known GPU raycasting algorithm has been employed to provide a 3D view of the colon. The goal of this module is to explore individual cases that caught the user’s attention, to better understand what are the values of the variables. Thus, the colon segments are rendered with an opaque transfer function with a different color for each segment, while the rest of the MRI light is semi-transparent to provide context. Since our goal was the development of a web application, raycasting has been implemented with WebGL, which is supported by the major browser vendors, but imposes some restrictions in terms of computational power and performance as explained next.

5.1. WebGL Volume Raycasting

Although graphics capabilities of current browsers have improved a lot in the recent years, they are still limited compared to the graphics capabilities available for standard desktop applications, where the full computational power of modern GPUs is exploited to obtain high quality visualizations in real-time.

Since the extensive use of interpolation and texture accesses have a great impact on performance depending on the device and the browser environment, we had to optimize our implementation to guarantee an interactive exploration. In our web
Fig. 6. Screenshot of our application showing the different parts of the Inspector Widget. (a) identifies the widget. (b) is a basic uni-dimensional transfer function editor: the user can change the color and the opacity of the density values by defining rectangular ranges. (c) shows the ray casting of the model, with each colon segment painted with a different color. The window identified as (d) provides the details on the calculated parameters for the patient. The identifier of the patient (anonymized) is shown at the bottom left part of the window, and the capture date is shown at the bottom right (illustrated as (e)).

application, medical datasets are loaded from DICOM files using the Daikon reader [26] and stored in a 3D texture containing the anatomical data in the positive range and the segmentation identifiers in the negative range. Then, rendering is performed using raycasting with early ray termination [27]. We experimented with different shading methods (see Figure 7) to find the proper balance between image quality and performance in our WebGL implementation: from just using a flat color for each colon component (a) to a basic emission-absorption method with diffuse shading (Phong model without specularity) (b). A cost-effective implementation of ambient occlusion (c) has also been implemented to improve the lighting quality (see Section 5.2). As it is shown in Figure 7(d), the combination of diffuse shading and ambient occlusion is the best choice to correctly convey the morphology of the colon. For this reason, we use this shading in our web application. Note that the staircase artifacts are due to the low resolution of the MRI images, and are not an artifact of the shading algorithm itself.

5.2. Optimizations

Ambient occlusion (AO) [28, 29] is an illumination technique that shades points as a function of the visibility from the environment. Points occluded by nearby geometry are darkened proportionally to the occlusion degree. The ambient occlusion technique that we have implemented is a modified version of Starcraft’s algorithm by Fillion and McNaughton [30]. The main difference is how we define the sampling hemisphere: we uniformly sample a voxel’s 18-connected neighborhood (the 6 face-adjacent voxels and the 12 edge-adjacent ones) as depicted in Figure 8. However, we only consider the sample directions that give non-negative dot products with the voxel’s normal. Thus, we roughly approximate the sampling along the surface of a hemisphere, which increases performance, and even sample directions perpendicular to the normal are taken into account. We tried using directions that only gave strictly positive dot products, which happen to work more correctly. However, the former method happens to highlight silhouettes along the colon segments which also gives some additional cues about the intestine’s morphology (see Figure 9).

5.3. Performance

In order to assess the performance and the memory requirements of the proposed application, we have carried out several tests using two different browsers (64-bit versions of Chromium 69.0.3497.92 and Firefox ESR 52.9.0) on a laptop with a Linux OS. The system is equipped with an Intel 64-bit quad-core i7-8550U processor operating at 1.8 GHz, 16 GB of RAM and a NVIDIA GeForce 930MX graphics card with 2 GB of memory. The resolution of the datasets used in this paper are $256 \times 256 \times 50$ and $232 \times 256 \times 50$ voxels. The application has been implemented using HTML, Javascript, and WebGL.

The setup of our application consists in generating the empty three large plots visualized in the website and every small plot at the bottom right cell (see Figure 3), setting up the entire inspector widget and its GPU ray caster behind the scenes. This process takes approximately 2.3 seconds (time mostly depends on the responsiveness of the server that hosts the website). We analyzed Chromium’s profiler statistics and memory consum-
The different shading techniques implemented: (a) the colon segments are identified by their color, assigned through the user interface, (b) shows the basic lighting with diffuse shading, (c) uses basic color and our simple implementation of ambient occlusion (see Section 5.2 for details), and (d) combines the diffuse shading and AO. The staircase artifacts are due to the small resolution in the Z dimension of the model ($256 \times 256 \times 50$).

Concerning the current status of the tool, the physicians found the 2D part as the most useful one, since it greatly facilitated finding correlations. Also the interaction was found to be simple and intuitive. Improving the 3D part was not considered as high priority, since it is good enough to provide information on the geometry of the different segments.

Medical doctors were very interested in extending the application so that it could also be used to compare the colon content data with data extracted with other means. Concretely, they were interested in the analysis of correlation between contents and the colon motility that can be obtained through a wireless video capsule capturing device, with a process named wireless video capsule endoscopy. However, such variables cannot be added straightforwardly, since this case does not have symmetry for the comparisons. Thus, we changed the SPLOM layout to fit all the desired variables. Since the number of motility variables is higher than the morphology ones, we changed the arrangement to a rectangular one with a larger aspect ratio, and thus, had to drop one of the larger scatterplots. Apart from this, we introduced other visual cues to ease the reading of the values in the matrix, as we explain in the following Section.

7. Motility data analysis

Domain experts suggested incorporating new data to our tool. They were currently analyzing the effect of cystic fibrosis in the motility of the colon and wanted to see how our approach might facilitate the analysis. In this section, we first detail how the motility data is created, and then, how we modify our tool to facilitate its visual analysis.

7.1. Motility database construction

The data of the second experiment was gathered by our collaborators using our framework and the capsule software, as described below. This database has been used for a publication that is currently under review. We asked the experts to use the same dataset with the exploratory visualization module. The goal of the experiment was to evaluate whether they would be
able to do the same analysis required for their research, and if it would be faster with our tool.

Besides the MRI scans acquired for all the patients of group B (described in Section 3.3.1) on separated days, intraluminal images (also referred as endoluminal images) of the gut were obtained by capsule endoscopy.

7.1.1. Acquisition protocol

Endoluminal images were obtained with a Pillcam SB2 video capsule and analyzed with the software by the same vendor: Given Imaging, Yokneam, Israel. Two images per second were obtained at a fixed rate and recorded during a total of 8 hours, with the subjects lying comfortably on a hospital bed and the trunk raised 30° above horizontal. Gastric exit of the capsule was determined by visual inspection at ten minutes intervals using a real-time viewer monitor. Forty-five minutes later, participants were instructed to ingest a liquid meal. After the eight hours procedure, images were transferred to the standard viewer program. After visual detection of gastric exit and cecal arrival, small bowel images were selected and examined to detect morphological mucosal abnormalities by an experienced gastroenterologist. Endoluminal images of the small bowel were analyzed with a computer program specifically developed in their laboratory for the evaluation of intestinal motility [31].

The image analysis software allows the detection and quantification of contractile and non-contractile patterns characterizing small bowel motility. Moreover the motion of both intestinal walls and content are also measured to detect low-motion and high-motion sequences characterized by established threshold. All these parameters were also quantified in different time periods (called phases), which are established by the medical experts.

7.1.2. Data extraction procedure

For this clinical study, a total of 16 (group B) participants were collected from adult patients with diagnosis of cystic fibrosis (CF). The goal of the study was to find if the delayed gut transit and pooling of contents in CF patients could be also due to an impaired intestinal motor function. In previous studies, this effect was related to an impaired regulation of a digestive secretory function. A small bowel motility evaluated with the endoluminal images could be complementarily confirmed with increased ileal and colonic volumes related to pooling of content. The objective of this clinical study was to search for significant correlations between the abnormalities detected by intraluminal (internal) images and the morpho-volumetric analysis of MRI scans (external images). Moreover, our first database of healthy subjects (group A) described in Section 3.3.2 was also incorporated to this study in order to have a reference of the CF patients (group B) in the MRI analysis. The statistical analysis in order to determine whether there was a significant difference of the mean with respect to each variable consisted of a Student’s unpaired t-test or a Mann-Whitney U test depending on the distribution of the data.

The endoluminal image analysis software generates a set of 91 variables. However, in our framework, we have only used the 24 variables deemed most important by medical experts. These are:

- Percentage of time the bowel segment is in contraction.
- Percentage of time the bowel segment is not in contraction.
- Percentage of time the image shows turbid moving contents (colon walls are not visible).
- Percentage of time the image shows static turbid contents.

Each of them is characterized in the six different phases established by the medical experts:

- All: includes the whole transit of the capsule.
- Pre: before food ingestion.
- Post 1h: in the first hour after the food ingestion.
- Post 1st, 2nd, and 3rd third: after food ingestion, along the first, second, and last third of the measured time.

7.2. Visual analysis of colonic content and motility data

The physicians proposed the modification of the visualization tool to look for correlations between colon content data and the motility of the colon. The captured data was first analyzed by physicians with SPSS, which took them around five to six hours, and they were interested to know whether the same analysis would be possible (and faster) with our tool.

To visualize potential relationships with this new data, we needed to modify the 2D tool because in this case we were not
comparing variables of the same set. However, since the tool was found to be very useful and intuitive in our previous informal evaluation, we kept the design as much as possible. We also added statistical data and allowed filtering variables to facilitate the exploratory analysis. The hypothesis of our domain experts was that this tool could substitute their SPSS work if we added some statistical data they obtained through their analysis. Thus, we added the Pearson coefficient $r$, and the probability $p$, which now appear on hovering. Moreover, we also provided a simple feature for variables filtering through double click on the variables and a button to refresh the changes to the view.

For the evaluation, physicians selected a set of variables from the capsule dataset (24 variables), as well as the potentially interesting variables from the colon content (12 variables). To maximize the utility of the application, we plot all the variables at once, which results in a 2:1 aspect ratio scatterplot matrix. This forces us to reduce the size of the large individual charts, and remove one of them. The resulting design is shown in Figure 11. We added other information that was deemed important, such as the protocol of the MRI test, and the number of patients present in the visualization.

The interaction is the same as for the other 2D visual analysis module: the user can drag any of the desired plots of the scatterplot matrix to the top charts. The scatterplot matrix then illustrates the selected plots with two different colors. Again, details-on-demand appear upon hovering the small plots (variables compared) or the larger plots (individual patient information). We also increased the information available with a legend of the encoded Pearson coefficients, as suggested by the physicians. Moreover, the inspector widget has been remodelled in order to show all the variables gathered from the MRI and the capsule acquisitions (see Figure 12).

8. Evaluation

Once the extensions were added to facilitate the comparison of morphological data with motility data, we carried out a formal user study with two digestologists of two different hospitals, with fourteen and ten years of experience, respectively. The domain experts are not co-authors of the paper.

The procedure was the following: we created a web application that automatically loaded the colonic content data and the motility data extracted from the wireless video capsule device, and it was executed in a Chrome browser. The physicians then had to spend some minutes exploring the data and then answer a usability questionnaire. The experts had to answer the questions in a Likert scale where 1 means totally disagree, and 7 means totally agree. Besides, they had the opportunity to fill in an observations field to add some details if needed. The evaluation produced the following results:

- The web application is easy to understand: 5.5.
- It easily shows the correlation between variables: 6.5.
- It is easy to select the parameters to compare: 6.5.
- It is useful to inspect the data: 6.5.
- I would use this application in my daily work: 7.
- It is easier to analyze correlations with this tool than with other tools: 7.
- It is faster to obtain results with this tool than with other tools: 7.

From the comments on their experience, we learnt that both physicians were happy with our tool, and they appreciated the features. For instance, they commented that "though we are used to find correlations with many variables in SPSS, it is highly time consuming and the correlations are difficult to see and relate to each other". They also think that our tool might overcome the limitations exhibited by current software ("It is useful, especially [...] when we have a large amount of variables that are difficult to relate to each other").

Besides, physicians also suggested some improvements: the inclusion of other information such as the statistical significance of the data, and highlighting tools to emphasize charts within a certain correlation range. These improvements, as it has been explained in Section 7.2 have already been incorporated. Another extension mentioned by one expert was the possibility of including other data. This requires an input module that seems not difficult to add.

After the improvements, we had another follow-up meeting to show the experts our recent advances and they suggested to add some other statistical descriptors (such as PCA) to our tool. This would let them drop completely SPSS or Excel from their workflow and just perform the whole analysis in the same tool. They also believe that the current design allows them to directly get snapshots from the tool to their publication for data presentation.
Fig. 11. Modified 2D visualization tool for the exploration of colonic content vs motility data. Hovering over a plot shows the parameters with a label and the encoded Pearson coefficients. Also, the plots visualized in the larger versions are highlighted. Moreover, if some variables have been hidden, the current Pearson coefficient range is shown in green. By double-clicking in the label of the different variables, they can be hidden in order to allow the physicians to focus on a subgroup of the overall variables set.

We already discussed the number of samples in Section 4.2, but not how the number of variables scale. In the last experiment, the total number of variables required for the analysis was very large: 91 coming from the endoscopy system, 55 from the MRI analysis, and 24 from different questionnaires where symptoms were also quantified. However, we only used 24 from the endoscopy system and 12 from the MRI analysis in our plots, as indicated by the physicians. If the visual exploration of the whole dataset was necessary, the system could discard lower correlation charts as it is done in the first version of our prototype. As well, interactive filtering could be performed as in the motility analysis version. Should the number of variables be too large, we could use abstracted versions of the charts with glyph SPLOMs as initial overview, in the same way as it is done by Yates et al. [21]. But this is something that depends on the concrete experiment and thus, needs further discussion with the experts.

9. Conclusions and future work

In this paper we have developed a framework for the exploratory data analysis of morphological data and contents of the colon. To the authors’ knowledge, no previous system has been developed for inspecting this kind of data of the digestive system. Previous research has only dealt with parts of the colon, and requires much more manual effort to get measurements. Our system here has three novel components: first, an automatic morphological data extraction module (MGeom), to extract relevant features of the colon. Second, the framework, that allows experts to gather both colon content and morphology data from MRI images. This framework has been already successfully used by domain experts to better understand the defecation process, or the effect of diets in the intestine [25]. Finally, we created a 2D visual analysis tool (MVis2D) by adapting well-known visualization techniques, such as SPLOM, for our needs in this project. And we extended this last tool to deal with other external data that measures colon motility. The basic 2D tool focuses on finding correlations between data, and saves time for the domain experts. The main goal of the 2D component is to facilitate exploratory visualization of complex data. This is the reason why all the features of the visual inspection are built around a matrix structure of a SPLOM, which serves as an overview of the variable exploration space and gives hints on the data clusters that lie within. The new filtering tool also facilitates the navigation by hiding variables or groups of variables at once. As concluded by our collaborators the tool
greatly facilitates finding correlations within the data set effortlessly. Physicians believe that, with some additions regarding some statistics evaluation, it could save them several hours in analysis. We also added a 3D viewer as a details-on-demand widget (referred to as $M_{V(3)D}$). This component corresponds to the individual patient exploration tool. Its primary role is to show details-on-demand: we display the segmented colon and show patient measurements directly with a table. Physicians need to explore these data to better understand certain values in the charts. Since the application is web-based, showing two patients side by side is simple, we only need to instantiate the same viewer component twice.

After another follow-up discussion with our collaborators, we see that the current system can benefit from some extensions: first, we have designed the 2D SPLOMs to hold a subset of all the available variables, guided by the problem and under the instructions of medical experts. However, the original datasets contain more variables. Should we need to display all of them, some modifications would be needed. As already mentioned, in the case of a huge set, we could use glyph SPLOMs as in the approach by Yates et al. [21]. Second, in order to completely substitute some tools (e.g. Excel and SPSS) from the current physicians workflow, we would need to add some extra statistical calculations such as PCA. Finally, concerning the 3D module, DVR is performed with a GPU-based ray caster, which is accompanied by a custom transfer function widget that allows for color classification of the volume’s samples. This module might be enhanced with interactive tools to allow the user measuring on demand different features of the displayed structures.

Finally, we would also like to present and extend our prototype to other kinds of data in the medical domain. For Gastroenterology, this might include other data on patients’ symptoms or conditions, for example. Extending the tool to other medical domains where they perform a related data analysis might also be useful for a broader potential audience as well as users.

**Acknowledgments**

The authors want to thank the medical experts that have helped us with the user study, F. Azpiroz and C. Malagelada from the Vall d’Hebron Hospital and A. Bendezá from the General Hospital of Catalunya. We also want to thank the anonymous reviewers that, throughout the reviewing process, proposed valuable comments that helped to improve the paper. This project has been supported by the project TIN2017-88515-C2-1-R (GEN3DLIVE), from the Spanish Ministerio de Economía y Competitividad with FEDER (EU) funds.

**References**


