Integrated Visual Analysis for Heterogeneous Datasets in Cohort Studies

Martijn D. Steenwijk, Julien Milles, Mark A. van Buchem,
Johan H.C. Reiber, *Fellow, IEEE* and Charl P. Botha, *Member, IEEE*

Abstract—Current medical research is often hypothesis driven, focusing on a limited number of parameters showing, or expected to show, some relation with the disease. When a supporting scientific ground or proper hypothesis is lacking however, this approach is not always fruitful. Visual analytics has seen limited application in medical research. We propose that visual analytics can be used to study parameters across patients, especially in cases where no clear hypothesis is available from the start. This can help medical researchers to focus their efforts. We present a visual analysis framework which provides highly interactive visual analysis of cohort data, and is able to deal with irregular multi-timepoint, imaging and non-imaging data. The framework integrates the extraction of features into the process of visual analysis and makes use of a carefully designed data structure able to keep track of data dependencies and interrelationships in inhomogeneous cohort study data. We evaluated the framework on a cohort of patients suspected of having neuropsychiatric SLE, a heterogeneous rheumatic disease. Visual analysis revealed a number of observations corroborating earlier findings. We were also able to identify new trends in the data that could indicate directions for further research, and illustrated thereby the potential of visual analytics to operate as a hypothesis generating tool.

Index Terms—Visual analytics, healthcare, medical visualization, medical cohorts, multi-modal, multi-timepoint.

1 Introduction

Current medical research is often hypothesis driven, focusing on a limited number of parameters showing, or expected to show, some relation with the disease. In current medical research it is common to analyze only a few measurements at the same time. When a supporting scientific ground or proper hypothesis is lacking however, this approach is not always fruitful. Visual analytics has shown to be of use in the analysis of large multi-parameter and multi-timepoint datasets, such as meteorological models. Visual analytics has seen limited application in medical research, primarily for studying inter-patient phenomena. From this research, it appears that visual analytics can be used as a hypothesis generating tool.

We propose that visual analytics can also be used to study parameters across patients. This might be particularly useful in cases where no clear hypothesis is available from the start. This could help medical researchers to focus on the proper parameters. In this paper, we propose a visual analytics framework and associated data structure that helps the user to extract parameters of interest, visualize parameters of interest, provides direct feedback and allows for easy exploration of multi-modal and multi-timepoint parameters across patients. The contributions of this work can be summarized as:

- Integration of a full medical cohort analysis work flow in a reusable and extensible implementation framework;
- Highly interactive visual and basic statistical analysis of cohort data, dealing with irregular multi-timepoint, mixed imaging and non-imaging, and missing data;
- Integration of online distributed feature computation in the visual analysis process;
- Design of a data structure and corresponding relational database layout, that enable keeping track of the data dependencies and interrelationships in inhomogeneous cohort study data, and easy extensibility in terms of patients and data types.

- *M.D. Steenwijk, and C.P. Botha are with the Computer Graphics section, Dept. of Mediamatics, Delft University of Technology.*
- *M.D. Steenwijk, J. Milles, J.H.C. Reiber and C.P. Botha are with the Div. of Image Processing, Leiden University Medical Center.*
- *M.A. van Buchem is with the Dept. of Radiology, Leiden University Medical Center.*

We have implemented the framework documented in this paper in a software system called PAPILIO, which consists of a preprocessing environment called PREPAP and a visual exploration environment called VISPAP. We evaluated the usability of visual analytics in the analysis of heterogeneous cohort datasets by applying PAPILIO on a cohort of patients suspected of the heterogeneous rheumatic disease neuropsychiatric systemic erythematosus (NPSLE). The framework is however applicable and scalable to other diseases and might be of use in the analysis of population-based studies and electronic health records.

The paper is structured as follows: related work is discussed in Section 2 and in Section 3 we describe our approach in detail. Section 4 addresses the details of our implementation and in Section 5 we evaluate the our work by performing an informal case study. Finally, we present conclusions and future work in Section 6.

2 Related work

Related work concerning visual exploration in medical datasets can be divided in two varieties: the study of medical data originating from a single patient and the study of medical data originating from multiple patients. This section is concerned with interactive visualization techniques that facilitate the finding of meaningful patterns in both varieties.

Visual data analysis combining multi-dimensional single patient medical data with real-time linked scatter plots and parallel coordinates representations was first investigated in the WEAVE system of Gresh *et al.* [14]. A number of aspects with regards to the specification of this kind of focus+context visualization were formalized by Doleisch *et al.* [8]. In 2007, Blaas *et al.* adopted this approach and added feature derivation possibilities, pattern recognition techniques and interactive multi-dimensional segmentation strategies [5]. So far, the approach was solely applied on single timepoint data. Oeltze *et al.* extended the application domain however to dynamic 3D perfusion data by deriving parameters from the time intensity curves in the dataset [18]. Fang *et al.* used another approach of computing similarities between spatial measures and time intensity curves to visualize dynamic 3D medical data [13]. Keefe *et al.* used small multiples to visualize biomechanical motion data [17]. In all these cases, the focus was on studying the behavior of medical data in a single patient or subject.

Research on interactive visual exploration of multi-patient datasets is scarce, usually focuses on the analysis of events or scalar data over time and is commonly found in the application of analyzing electronic

health records (EHRs). Lifelines [19] present personal history record data organized in expandable facets, but does provide – aside from panning, semantic zooming and text filters – very few ways to manipulate the data. PatternFinder [12] presents a form-based query interface for specifying temporal queries over patient histories. These forms give the use extensive control in filtering, but they are also very complex to specify. In 2008, Wang *et al.* presented the alignment, rank and filter (ARF) framework which firstly enables the user to interactively align time-oriented data across patients [24]. This approach was later extended with categorical searching, aggregation and group comparison [25]. In all these cases, imaging data was however not considered.

A conceptual framework integrating the visual analysis of multimodal and multi-timepoint data across patients was presented by Blaas in 2010 [4]. Departing from the concept of multi-field data, Blaas defined a framework of domains, features and mappers which allowed visual exploration of image data across multiple medical imaging modalities at the same time, with a link between views on patient, group and voxel level. From the evaluation on a dataset consisting of patients with multiple sclerosis, it appeared that the conceptual framework could be used as hypothesis generating framework. Due to its promising results, we adopted Blaas' conceptual framework as a starting point for our work.

3 METHODS

At its core, our framework makes it possible to combine various types of raw medical data into a searchable database, fully integrated with an interactive data exploration environment. In cohort studies, imaging and/or non-imaging data is usually available for multiple patients, can originate from multiple modalities and might be available at multiple timepoints. Usually, this data is unordered and stored in different locations. Our approach is to preprocess and store the imaging and non-imaging data in a searchable database. The layout of this database is designed in such a way that it enables the flexible application of visual exploration techniques. The visual exploration techniques visualize the data with different representations and initiate dynamic search queries based on the user's interaction. Since all data dependencies are stored in the database, this enables the user to interactively explore relations in the data.

A carefully designed and implementable data structure is required to integrate data processing and interactive visualization of multimodal and multi-timepoint data in one framework. Therefore we propose the use of domains, features, mappers and studies (see Section 3.1) which will organize the data such that imaging and non-imaging data can be stored and explored. After importing this data, careful pre-processing is usually required to prepare it for analysis. Therefore we propose to use a visual preprocessing environment in which the preprocessing pipelines are displayed as diagrams (see Section 3.2). The preprocessing environment's back-end takes care of the data organization, storage in the database and feature extraction for the available datasets. The data stored in the database can subsequently be explored using various visual exploration techniques, which initiate dynamic database queries and are discussed in Section 3.3).

3.1 Conceptual framework

The framework we propose is based on four concepts, including *domains*, *features*, *mappers* as inspired by Blaas [4] as well as our newly added concept of *studies* (see Fig. 1). In short, a domain defines the structure of a space, features define the values on that space, mappers define the relation between points on two domains and studies connect features measured in the same patient and on the same day but originating from different domains. The following sections explain these concepts in detail.

3.1.1 Domains

A domain defines the structure of a space on which features are defined and can be measured. In medical imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI), the domain is usually defined as a 3D spatial grid. In laboratory testing however, test results are usually defined as a scalar value. For example, the erythrocyte sedimentation rate (ESR) is defined on the ESR domain: a domain spanning all valid ESR measurements. Other domains include for example neuropsychological test results, or intermediate representations in the analysis of medical imaging data, such as histograms. The type and properties of a domain determine which operations can be applied on its features. This means for example in the case of 2D image processing algorithms, that they can be applied on any other dataset where a 2D structure exist.

We distinguish *normalized* and *non-normalized* domains. Normalized domains contain features that can be compared across patients, whereas non-normalized domains can be used to derive normalized domains from. A typical normalization procedure is defined by a combination of mappers and aims to derive meaningful features that can be compaired across patients. For example, raw MRI scans of individual patients are considered to be non-normalized domains. The scans are not registered and cannot be used to compare between patients. A typical normalization procedure would compute a normalized quantitative parameter from the scan, or register the scans from different patients to a shared spatial reference frame. These normalized features can subsequently be used to compare between patients.

3.1.2 Features

Features define the actual values on a domain. Features can be measured (as in CT) or computed from other features using mappers. A feature can also be seen as a function that maps a point in a domain to one or multiple data values, depending on the structure of the domain. In the case of a MRI scan for example, the signal intensities in the data volume are regarded as being a feature. Here the feature takes the form of a 3D dataset, since its domain defines that MRI scans consist of 3D datasets. In the case of a laboratory test, the result of the test is regarded as a feature. In this case, the feature will often take the form of a single value. Features might also include qualitative measures (e.g. "positive" or "negative" test result). Our implementation framework stores the data values of features accompanied with their patient IDs and examination dates. This way, the components of a study can be easily determined.

3.1.3 Mappers

A mapper defines the relation between features on different domains. Mappers are for example used in for deriving normalized features from non-normalized image data. For instance, an algorithm performing a "brain extraction" from a T1-weighted MRI scan is regarded as a mapper which maps the features from the domain "T1 scan" to a new domain called "brain extracted T1 scan". In other words, this mapper extracts the brain from the loaded T1 scan and stores the results as features on the domain "brain extracted T1 scan".

Although mappers transfer features from one domain to another domain, they have the property of preserving dependencies and relationships in the data. In this way, the user can always drill down from a derived feature to the original data.

3.1.4 Studies

The goal of the "study" concept is to combine related features from different domains and reflects the state of a patient at a certain moment. A study consists of all features measured in a patient on a certain date and is defined by a patient's ID and an examination date. For instance, a patient could have had a MRI examination and a laboratory test on the same date. The resulting MRI scan (and its derived features) and the laboratory test result are considered as part of the same study.

Conceptually, a study can also be interpreted as a point having n features on n domains. The n dimensional point defines a state of a patient in the n dimensional space of patient states. This space is composed of the n domains currently included in the framework. The point consisting of the reduced number of normalized dimensions can be compared with other normalized patient states.

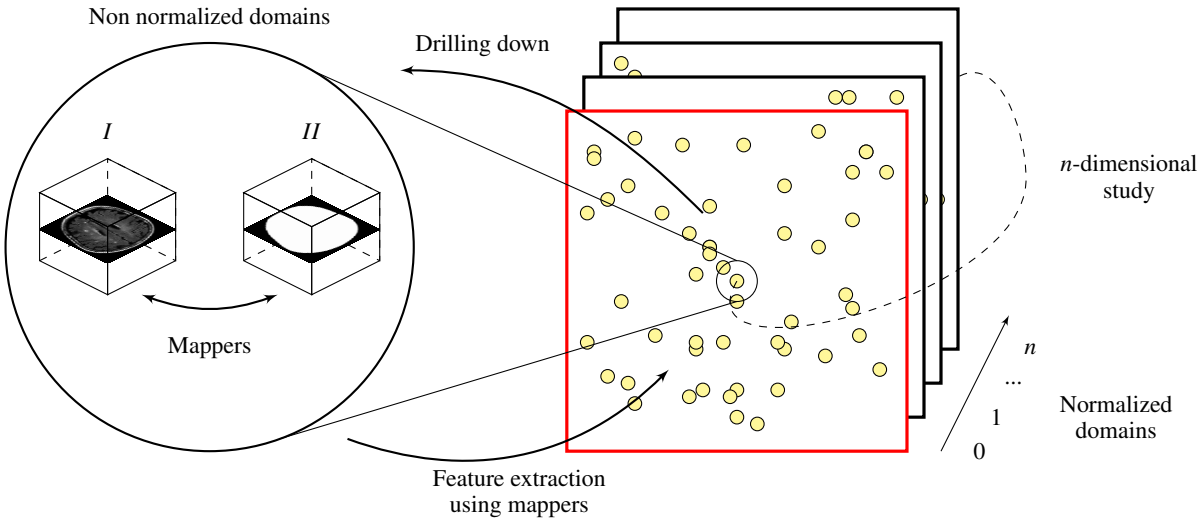


Fig. 1. A schematic representation of the conceptual framework and data structure. Domains I and II are non-normalized domains. The relation between domain I and II is given by mappers. Feature extraction is a process of mapping a domain to a normalized domain. A study is a point in the space of n normalized domains, connected by its patient ID and examination date.

3.2 Image preprocessing

Current medical research evaluates imaging data commonly qualitatively or quantitatively. Especially in the field of neuroimaging, where quantitative imaging techniques such as quantitative MRI (qMRI) have become available, data analysis is shifting more often to automated quantitative analysis techniques. This type of analysis is usually performed by applying a number of image processing algorithms on each dataset and aims to obtain a meaningful set of normalized parameters that are comparable between patients. These image processing algorithms often include unsupervised command-line tools, such as the FMRIB Software Library (FSL) [20]. The process of designing a batch script is not always intuitive for researchers with a medical background. Furthermore, the execution and optimization of a feature extraction pipeline for multiple patients is an error-prone process.

Therefore we supply a graphical boxes-and-lines interface, called PREPAP, in which a feature extraction pipeline can be visually designed using mappers supporting existing image processing tools. Since the mappers can be implemented as black boxes, the user does not require any programming experience to extract the features. Furthermore, this approach gives rise to several advantages:

- The diagram representation allows for intuitive design and customization of feature extraction pipelines;
- Mappers can be implemented such that the use of different processing libraries is supported;
- The diagram representation helps to ensure that all datasets are processed consistently;
- The graph-based representation enables additional functionality such as caching of intermediate results, scheduling and parallel preprocessing.

3.3 Visual exploration

For the visual analysis of the cohort data, we have created a multiple coordinated view interface, called VISPAP, where the views and other interface elements have been implemented as dockable windows. Fig. 2b shows the main interface with two docked views. The views display (normalized) scalar data or visualize volumetric image data and can be used in any number and combination. The representations are interconnected to support various interactive exploration techniques. This section discusses both the visual representations and interaction techniques.

3.3.1 Visual representations

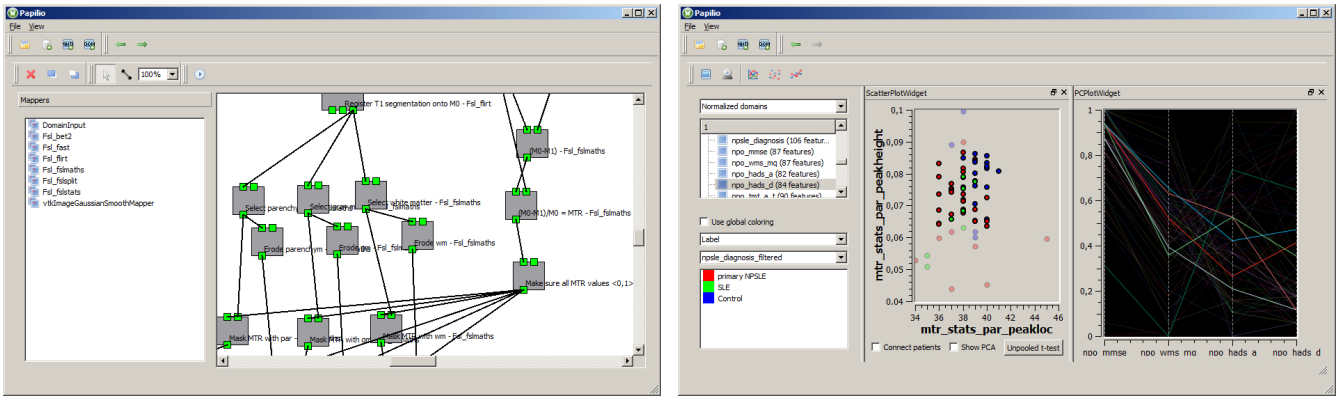
We distinguish two types of visual representations. The first type consist of plots in which (high-dimensional) feature data can be visualized. The second type includes visualizations in which volume data can be visualized. High-dimensional feature data is visualized using scatter plots, parallel coordinates plots (PCPs) and time plots. Data is loaded into these representations by dragging individual normalized domains, from a list in the interface, onto the axes of the individual representation.

The *scatter plot* representation visualizes the features on two domains and can be colored by the features of a third domain. In our application, each marker on a scatter plot represents a study (see Fig. 3(a)). To visualize the relation between studies originating from the same patient, markers originating from the same patient can be connected. The advantage of a scatter plot is that relations between or clusters within the features on two domains can be easily identified. A disadvantage of a standard scatter plot is the occurrence of visual clutter when visualizing a large number of feature points.

A common way to visualize high-dimensional data is by using a *parallel coordinates representation* [16]. In this representation, the n-dimensional feature space is visualized on a grid consisting of n parallel lines, where each point in the feature space is represented by a line intersecting the n parallel lines. Using this representation, the features on an arbitrary number of domains can be visualized in one visual representation. By altering the order of the axes, scaling and coloring, PCPs can be used to identify trends in high dimensional data. In our application, a line in the PCP represents a study consisting of n domains. This is illustrated in Fig. 3(c) and 3(d).

The behavior of a certain feature over time can be displayed using a *time plot*. This plot displays the time course for a certain feature in a patient. Inspired by the ARF-framework of Wang *et al.* [24], we organize the longitudinal data in this plot by aligning the time courses by date, by patient's birth date or by first occurrence of the feature (see Fig. 4). Each alignment has its own advantages. Alignment by date visualizes when the feature was measured in the cohort, alignment by patient's birth visualizes the trend of a feature at a certain age and alignment by first occurrence displays follow up of individual patients.

Finally, a 2D slicer and 3D renderer can be used to visualize volumetric image data, for example when the user drills down on a specific marker in a scatter plot. These visualizations are particularly useful during the inspection of source image data, and during checking the functionality of the feature extraction pipeline.



(a) (b)

Fig. 2. PAPILIO's user interface. Fig. (a) shows the PREPAP feature extraction environment. On the left a list containing the available mappers. These are displayed as blocks in the feature extraction pipeline. On double-clicking a mapper, mapper parameters can be altered. Fig. (b) shows the VISPAP visual exploration environment. On the left top, a list of available normalized domains is displayed. Studies are visualized in the representations at the right, which can be colored using the coloring panel at the left bottom.

3.3.2 Interaction techniques

We provide a number of basic interaction techniques to allow users to directly interact with the visualizations and interactively query the data in the database. The current interaction methods include linked selection, coloring, drilling down and statistical-aids.

Linked selection is an interactive method to explore high-dimensional data [8]. The selection of studies in one representation, will highlight the corresponding studies in other representations. This way, the behavior of cluster of studies in one representation can be visually explored in other representations.

Coloring can be used for different purposes. First, studies originating from the same patient can be colored with the same color. This gives an impression of the behavior of features within a single patient. Studies can also be colored by label. A 'label' is a qualitative feature, for instance a diagnosis or the presence of a certain antibody in laboratory testing. On coloring by label, groups of studies with a common aspect are visualized. This way, the spread of a diagnosis or effect of the presence of a certain parameter can be explored. Finally, studies can be colored by a domain containing quantitative features. By mapping the feature values on a color scale, the effect of the change of a certain parameter can be studied in the different plot representations.

We use *drilling down* in two ways. First, a user might want to select an interesting subset of the data to focus on. This can be done by selecting this subset and removing the non-selected studies from the representation(s). Second, drilling down can be used to verify the integrity of the source or preprocessed image data. If a scatter plot visualizing image derived data for instance displays an outlier, the outlier can be drilled down to. By right-clicking the outlier, all parent features (including image data) can be directly accessed and visualized in the 2D slicer or 3D renderer.

Finally we provide the possibility to make use of basic statistical analysis methods when coloring by label in the scatter plot. The global distribution of the various groups can be visualized by displaying optionally confidence-weighted principal component ellipses of the individual groups. As can be seen in Fig. 5 this helps to understand the behaviour of a complete group of patients by visually aggregating their data. Furthermore, the significance of the difference between the individual groups can be easily computed.

4 IMPLEMENTATION FRAMEWORK

The conceptual framework has been implemented in a C++ software system called PAPILIO. We used Trolltech's Qt for the user interface. The image processing and registration components make use of the FMRIB Software Library (FSL), while plot components are based on the QwtPlot library and visualization components are based on the Visualization Toolkit (VTK). We chose SQLite as the central database management system because of its light weight and serverless operation.

The user interface (see Fig. 2) consists of two tabbed environments. The first environment called PREPAP takes care of importing and preprocessing the data. The second environment is called VISPAP and provides an interface for visual exploration. RAW data can be imported from various formats using the various LOADER modules provided in PREPAP. Supported formats include imaging data such as DICOM, VTI and NIfTI, but also non-imaging database formats such as CSV and the frequently used SPSS and Microsoft Access database formats. Features can be derived from image data using PREPAP's feature extraction diagram. These can be subsequently visually explored using linked scatter plots, PCPs and time plots in VISPAP.

PREPAP provides a toolbar with import modules at the top and a list with available mappers in the left panel. These mappers can be dragged onto the feature extraction canvas, and subsequently linked together by dragging lines between the in- and outputs. The current implementation includes mappers using FSL and VTK image processing algorithms, and has support for caching and distributed processing.

VISPAP provides a toolbar with the available representations at the top, and provides a list of the normalized domains and a coloring panel in the left panel. Representations are displayed in the right part of the screen. Domains can be dragged onto the axes of the various representations. This way, features can be easily loaded into the various representations. The coloring panel can be used to devise a global coloring mode or a coloring mode for individual representations.

5 INFORMAL CASE STUDY

We evaluated the usability of the framework by applying PAPILIO on a cohort of patients suspected of neuropsychiatric systemic lupus erythematosus (NPSLE), a rheumatic disease of the central nerve system with very heterogeneous symptoms. The heterogeneity and amount of available data (MRI, neuropsychologic test results and laboratory test results), as well as related work from literature, indicated that visual analysis could be helpful to generate higher confidence hypotheses than has been possible up to now. During the evaluation, we aimed to establish whether visual analytics could corroborate earlier findings and further whether it could help point out new hypotheses.

5.1 Neuropsychiatric SLE

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by various symptoms and the presence of various antibodies [9]. SLE is diagnosed based on complaints, symptoms and/or abnormal laboratory test results. Since symptoms are different for each patient, the American College of Rheumatology (ACR) defined 11 criteria to differentiate SLE patients in research [22]. Criteria are for ex-

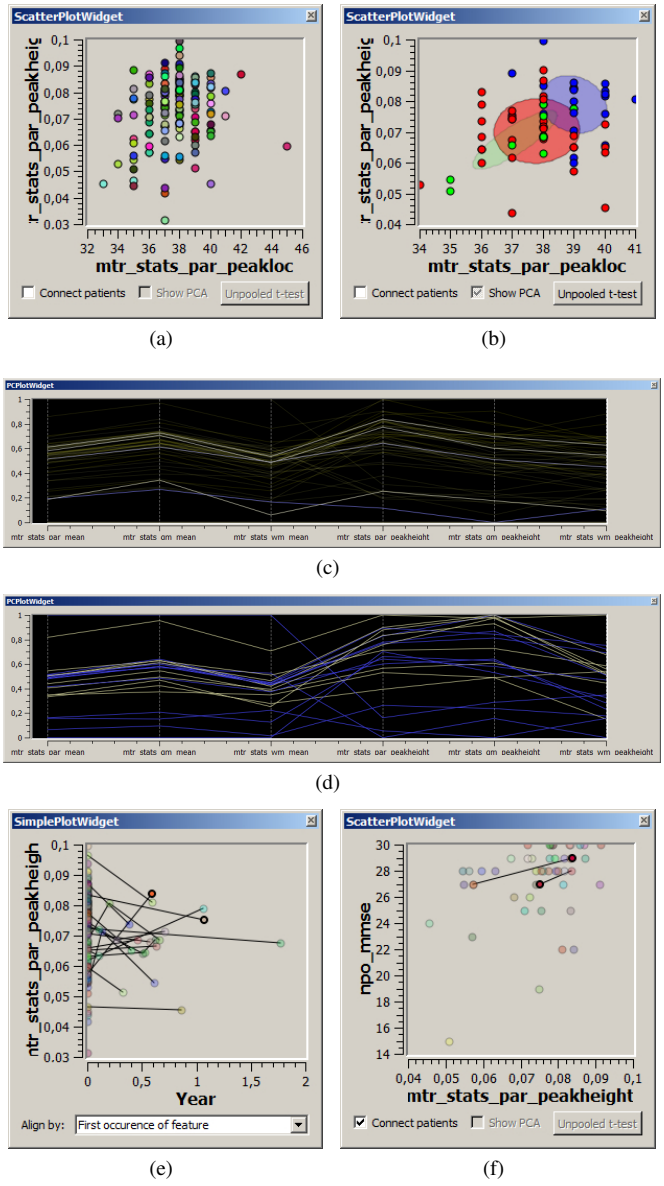


Fig. 3. Visual analysis of image data. Fig. (a) and (b) show whole MTR peak height versus location. Fig. (a) is colored by patient, and shows an outlier with a high peak location. The outlier is removed in (b), where the studies are colored by diagnosis (red: primary SLE; green: SLE; and blue: control) and shows a PCA ellipse for each group. Fig. (c) and (d) show various MTR parameters colored by MMSE (Blue to yellow color scale, respectively $\langle 15,30\rangle$ and $\langle 27,30\rangle$). Fig. (c) does not show a clear separation for the selected MMSE scores lower than 25. Narrowing the color scale in (d) shows a separation in especially peak height measures. Fig. (e) and (f) suggest that increase in whole brain MTR peak height might be related to an increase in MMSE in an individual patient.

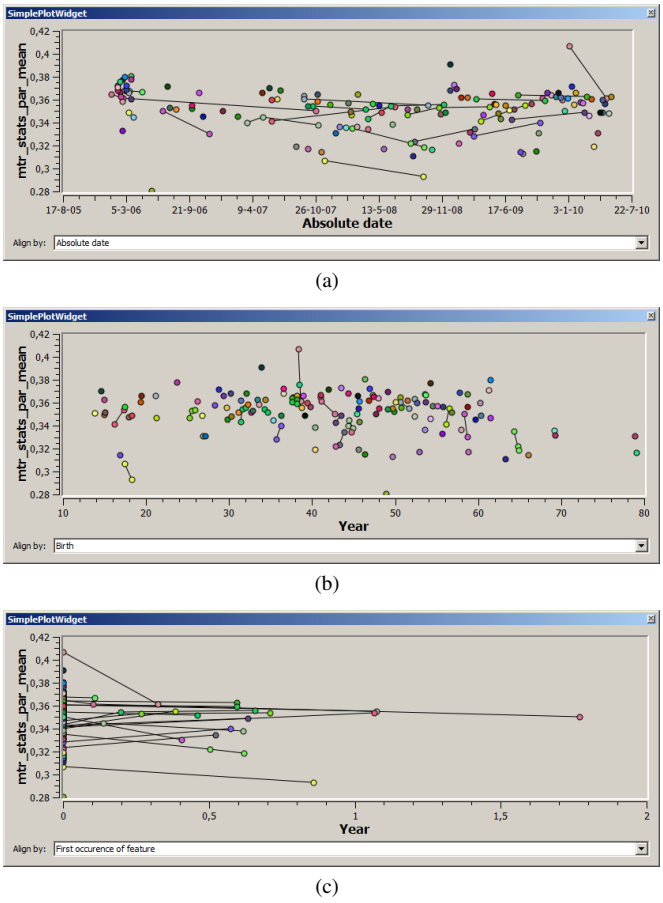
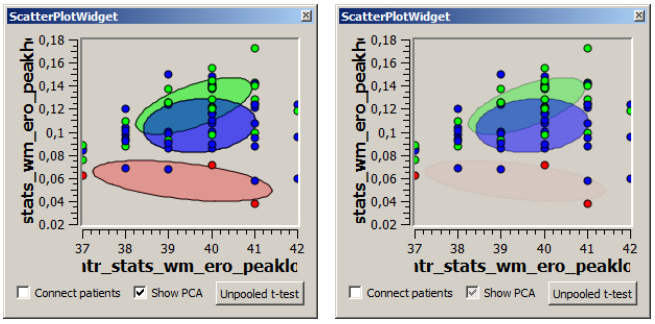


Fig. 4. Alignment of studies by (a) study date, (b) date of birth, and (c) first occurence.

ample the typical malar rash, discoid rash, hypersensitivity to sunlight, pericarditis, glomerulonefritis or the presence of certain antibodies. A patient has to fulfil four or more of the criteria to be classified as SLE patient in research. In practice, these criteria are also used to diagnose SLE in the clinic.

One of the ACR-criteria for SLE denotes "neuropsychiatric disorders". A large number of the SLE patients develop neurologic, psychiatric or cognitive symptoms during the course of the disease. Neuropsychiatric manifestations of systemic lupus erythematosus (NPSLE) vary from mild to severe and often are difficult to diagnose and distinguish from those of other diseases. Each part of the nervous system could be affected, which may cause symptoms ranging from cognitive dysfunction to seizures and strokes. Although the underlying pathophysiology of NPSLE is unknown, current literature suggests the involvement of different pathomechanisms including vasculitis, antibody mediated activation of the coagulation system and antineuronal antibody mediated immune responses [15]. NPSLE is diagnosed *per exclusionem* and the incidence ranges from 14% to 80% [6]. This large extent is caused by the method of measurement in different studies, difficulty of diagnosis and variety of symptoms. In order to help clinicians and researchers define NPSLE more clearly, the ACR also developed standardized case definitions for NPSLE [1]. This classification system not only describes diagnostic criteria, but also exclusion criteria.

The diagnosis of NPSLE often is a dilemma. Symptoms are nonspecific and diverse, furthermore a reliable diagnostic test is lacking. It is often unclear whether and how NPSLE patients should be treated. Some researchers believe that a difference between SLE and NPSLE does not even exists, but that some patients are simply more sensitive for the neuropsychiatric symptoms. Results of preliminary studies show that imaging techniques, in particular MRI, can be used to visualize brain damage. For example, magnetization transfer imaging (MTI) can help making a diagnosis of NPSLE in individual patients [10], and also demonstrated the ability to quantify the amount of brain damage [11]. Quantitative analysis also showed that invisible damage exists beside visible lesions. The pathophysiology of both is however unknown. Further research classifying and quantifying brain damage is crucial for improving diagnosis and treatment of NPSLE.



(a) (b)

Fig. 5. Principal component ellipses of three groups visualized in the scatter plot. The red ellipse in scatter plot (a) suggests a very large space spanned by the red colored studies. The red group consists however of only three studies. In scatter plot (b) confidence visualization is enabled, which weights the opacities of the ellipses by the number of studies in each group.

5.2 Materials

This evaluation of the framework is based on one of the worlds largest cohorts of NPSLE patients. The Leiden University Medical Center (LUMC) started recruiting this population in 1996 and documented all patients thoroughly. The cohort contains a large collection of 1.5T MRI data without a standardized protocol. As of 2004, patients were scanned on a 3T MRI with a standardized protocol, including a T1 and MTI sequence. This cohort, consisting of 154 patients suspected of having NPSLE, was used to evaluate PAPILIO.

Neuropsychological test results were available for a large subset of the patients. The neuropsychological tests were acquired using a standardized protocol, and included among others the hospital anxiety depression scales (HADS), the mini-mental state examination (MMSE), the Wechsler Memory Scale (WMS) and the trailmaking test (TMT). The tests were acquired on the day of the MRI examination, scores were normalized for age and level of education, and subsequently digitized by the department of neuropsychiatry.

Laboratory test results are at least available for all patients on the day of the MRI examination. For a subset of patients treated in the LUMC itself, all laboratory test results of the past 10 years were available. The laboratory testing protocol was not standardized, but included clinically relevant disease markers such as erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), anti-dsDNA and antiphospholipid (APL) antibodies in subsets of patients.

5.3 Data import and image preprocessing

The computation of meaningful MTI parameters requires a magnetization transfer images and anatomical, T1-weighted, scans. Here the MTI scans are used to compute magnetization transfer ratio's (MTRs), which provide quantitative information about tissue integrity. In healthy individuals, MTR histograms of the brain are characterized by the presence of a single, sharp peak, indicating that the brain is homogeneous in terms of MTR characteristics. MTR histograms are usually generated subvolumes of the brain, such as the whole brain or only gray matter. Subvolumes are obtained by segmenting the anatomical scan. To make the MTI and T1 scans available in PAPILIO, the available image data was imported from DICOM and NIfTI format. This resulted in 184 T1 examinations in 154 patients, and 160 MTI examinations in 134 patients. In total, 151 studies included both a T1 and MTI scan.

The neuropsychological test results were imported from SPSS. Subsequently, all laboratory test results of patients with imaging data were retrieved from the hospital information system. This resulted in a sample of over 75,000 records, of which the subset of interest was imported. Laboratory test results concerning the presence of antibodies were only available as qualitative measures and were imported as labels. Finally, the clinical diagnosis was available for 176 studies (148 patients). The diagnoses were divided in five main groups: including active primary NPSLE, residual primary SLE, secondary NPSLE and other diseases. Together with 20 controls, the diagnoses were imported in PAPILIO.

The MTR parameters were derived from the image data using the automatic preprocessing environment provided in PAPILIO. Based on an existing FSL script, a feature extraction diagram was designed which derived MTR features for whole brain, gray matter and white matter from T1 and MTI images. The features included segmentation volume (i.e. whole brain, gray matter and white matter brain volume), mean MTR value, MTR standard deviation, MTR peak height and MTR peak position. The preprocessing procedure derived these features for all 151 studies which had the complete source data available.

5.4 Visual analysis of image data

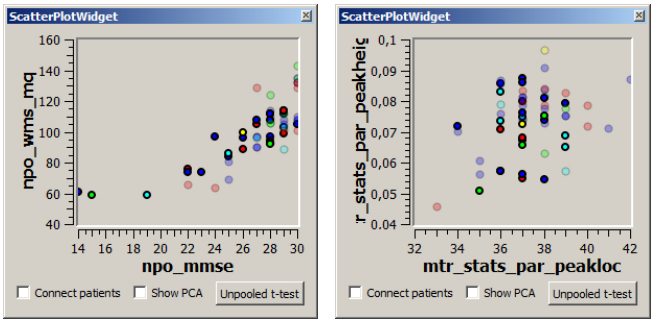
Initially we were interested in a quick assessment on whether the MTR parameters had been computed correctly. Therefore we opened a new scatter plot and dragged the domains whole brain MTR peak height and MTR peak position to the axes. Although the results seemed to be in the same range, an outlier could be identified of which the peak location was unusually high. By drilling down and visualizing the source MTI data in a 2D slicer, it was identified that the source data contained an acquisition artifact. Due to the acquisition artifact we removed this scan from further analysis by marking it as invalid (see Fig. 3(a)).

Previous research showed that patients with SLE and NPSLE have a lower MTR peak height than controls [11]. To investigate whether this could be supported by visual analysis of the current dataset, we colored the studies on the existing scatter plot by diagnosis and enabled the visualization principal component ellipses (see Fig. 3(b)). This representation indeed revealed a lower average peak height in patients with SLE and NPSLE compared to controls and corraborated this earlier finding.

MTR peak height has also been associated with cognitive dysfunction [11]. To investigate whether visual exploration supported this finding, we explored the relation between MTR parameters and the mini-mental state examination (MMSE) result[1]. Therefore we opened a new PCP and added the domains MTR mean value and MTR peak height for whole brain, gray matter and white matter to the plot. The studies in the PCP were colored by MMSE score. The studies having abnormal MMSE scores were selected by using a linked scatter plot. We expected this to highlight low MTR means and peak values in the PCP, but this was not observed (see Fig. 3(c)). Narrowing the color scale to normal MMSE scores (27 to 30) did reveal however a decrease in MMSE for reduced whole brain and gray matter MTR peak height (see Fig. 3(d)). This suggestion was confirmed using a scatter plot representation in which whole brain peak height was plotted against gray matter MTR peak height and the studies were colored by MMSE. Since the MMSE scores were corrected for age and education level, the observation in this small range can still be considered as of interest.

Another observation was done purely by coincidence. We used a simple plot to visualize whole brain MTR peak height and aligned the studies by first occurrence (see Fig. 3(e)). This representation clearly separates the patients with MTR followup from the ones without. By adding a new scatter plot displaying MMSE versus MTR peak height, and connecting the studies originating from the same patients, the change of MMSE and MTR peak height is displayed over time (see Fig. 3(f)). It was surprising that the trajectories of the two patients with follow up, almost have the same slope. From this plot, and by highlighting the corresponding studies in the time plot, it could be observed that in these two patients an increase in MTR peak height was accompanied by improved cognitive function, and vice versa.

[1]The mini-mental state examination is a neuropsychological test which investigates global cognitive function. MMSE scores range between 0 and 30, with values of 25 and above considered normal.



(a) (b)

Fig. 6. Two scatter plots showing MMSE, WMS_mq, MTR peak height and MTR peak location for all patients which have been tested on antiphospholipid antibodies. The selection shows the subset of patients with at least one antiphospholipid antibody present. The studies are colored by diagnosis: primary NPSLE (cyan), secondary NPSLE (blue), residual NPSLE (yellow), SLE (green) and other disease (red).

5.5 Exploring the role of antiphospholipid antibodies

Lupus anticoagulant (LAC), IgG anticardiolipin (IgG aCL) and IgM anticardiolipin (IgM aCL) are antiphospholipid antibodies. These antibodies are known to cause blood clotting problems, are often present in patients with NPSLE and are considered as a potential cause of neuropsychiatric symptoms since they might induce blood clotting problems in the brain on microscopic level. Furthermore, the presence of antiphospholipid antibodies is associated with cognitive dysfunction [7, 21].

Since MTR peak height is associated with cognitive dysfunction, it was interesting to investigate whether MTR peak height, MMSE score and the presence of antiphospholipid antibodies show any relationship. In conventional medical research, this would have been challenging to investigate since the data originated from three sources. By loading the APL antibodies from the hospital information system into our framework however, PAPILIO's data structure took care of the organization of the data which facilitated the investigation of the triumvirate.

We started the investigation by opening two scatter plots. In the first scatter plot, the domains MMSE and WMS_mq[2] were loaded. In the linked second scatter plot, the domains whole brain peak height and peak location were loaded. The studies in both scatter plots were colored by disease (see Fig. 6). Subsequently, a linked PCP was opened in which the three antiphospholipid antibodies were loaded. Using the PCP, all studies having at least one of the three APL antibodies positive were highlighted in both scatter plots. Unfortunately, no clear relation could be found in one of the scatter plots. Remarkably, what was found, was that all SLE and NPSLE patients in the cohort with abnormal cognitive dysfunction (MMSE < 25) had antiphospholipid antibodies present.

We also analyzed MTR peak height, MMSE and WMS_mq and the presence of antiphospholipid antibodies in a more general way. During this analysis, we loaded the neuropsychological tests and MTR parameters into scatter plots and colored them independent of the diagnosis by the presence of individual antiphospholipid antibodies. By making use of the integrated statistical hypothesis testing functionality, it appeared that in particular patients with positive LAC suffer from cognitive dysfunction and memory loss.

5.6 Discussion

Using PAPILIO, we demonstrated the usability of visual analysis techniques on medical cohort data. Imaging and non-imaging data from different modalities were included and analyzed longitudinally and across patients. The framework proved to be particularly helpful in the organization of the data, the automated analysis, the ease with which the data can be shown and the ease with which queries can be done.

In the application to NPSLE, visual analysis was used to explore MRI, neuropsychological and laboratory data. By drilling down we were able to identify invalid data and showed that visual exploration techniques can be applied to get quick insight into the data. Although the visual analysis was performed by non medical specialist, we were able to identify trends in the data which corroborate earlier findings. The potential of our system as a hypothesis generating framework was emphasized by the visual extraction of a number of interesting new relations that warrant additional research. For instance, linkage between a time-oriented representation and a scatter plot revealed that an increase in MTR peak height might be related to clinical improvement in cognitive function as measured by MMSE in an individual patient. Furthermore, it was found that all patients in the cohort with abnormal MMSE scores had APL antibodies present. Finally, a global analysis on antiphopholipid antibodies revealed that patients with LAC might suffer more often from memory complaints than patients without LAC. Were this hypothesis to be confirmed, NPSLE patients suffering from memory problems might possibly be helped by just prescribing "relatively simple" anticoagulation medication. It is important to note that our framework has helped to point out a number of potentially interesting relations, that need to be further investigated using traditional research methods.

5.7 Limitations

The above evaluation was performed by a non medical specialist. Although this person is knowledgable concerning neuropsychiatric SLE, the observations should be reconsidered by medical professionals to judge their interpretation. Furthermore, it should be noted that the diagnosis groups were devised by the same person. They should also be reconsidered by an experienced medical professional to enhance the validity of the results.

6 Conclusions and Future Work

In this work, we have shown that visual exploration across patients and multiple modalities could be a promising new approach in medical research. The technique provides researchers the opportunity to interactively gain new insights in their data, generate new hypotheses and exploit the potential of increasing amounts of medical data better. The combination of human analytical skills with computational techniques showed to have the potential to extract valuable information from medical data and might help medical researchers to focus their efforts.

At the basis of our hypothesis generating framework in which multivariate, multi-modal and multi-timepoint data can be visually explored across patients, is a carefully designed data structure consisting of *domains*, *features*, *mappers* and *studies*. This data structure organizes raw data such that it becomes suitable for further processing and visual exploration. The data structure is synchronized with a relational database. After importing, preprocessing and storing the cohort data into the database, we exploited the flexibility and speed of the database to present the data in the visual exploration interface using different representations and interaction methods.

We have evaluated the usability of our framework by applying it to a cohort of patients suspected of having NPSLE. Preliminary research indicated that visual explorations might be of use in this heterogeneous disease. By extracting features using the feature extraction diagram and visually exploring MRI, neuropsychological and laboratory data we were able to observe trends corroborating earlier findings. In addition, We observed a number of relations that warrant additional research. Although the value of our findings should be reconsidered by medical professionals, we believe that the current framework represents a promising first step in the application of visual analytics to medical cohort research.

In future versions, we plan to continue optimizing the various *visual representations*. For example, the performance of the scatter plot could be improved by providing a continuous scatter plot [3]

[2]The Wechsler Memory Scale is a neuropsychological test which investigates memory function. The average WMS mean quotient is 100.

and the PCP could be improved by using bi-directional frequency histograms [2].

During the evaluation, we identified that computer-aided exploration, by making use of dedicated data mining, statistical and pattern recognition techniques, would further strengthen the visual analysis process in the search for groups of features showing a similar behavior. A small preliminary study using PRTools [23] in MATLAB confirmed this, for example using a feature selection approach which selects the most corresponding domains to a certain domain, and can be used to suggest interesting feature combinations.

The evaluation also revealed the need for improved techniques to deal with *missing data*. Future work should develop improved techniques for detection and handling missing data. Here, considerable attention should be given to communicating the presence of incomplete data to the user, since data is often missing due to practical reasons (i.e. the data is available, but not as offline file, et cetera). It has also been identified that the current definition of a study should be made more flexible for practical use. Currently, the time slot in which features are combined to a study is only one day while certain measurements might be valid for a longer period of time.

Privacy protection. An unwanted side-effect of PAPILIO's approach to combine data from various modalities is that researchers might recognize individual patients based on the data. We plan to address this by integrating anonymization techniques into the framework. For instance hashing techniques might be used to anonymize patient ID's (e.g. MD5 would compute a unique hash for each patient ID) and defacing techniques to guarantee that image data is unrecognizable. The framework's structured design makes integration of this type of techniques relatively easy.

Finally, we will continue collaborating with medical researchers in order to further refine the functionality in our framework, and, importantly, to evaluate and judge the relationships that we have discovered in the cohort data up to now.

ACKNOWLEDGMENTS

The authors wish to thank Stephanie Greeve, Jeroen van der Grond, Jasper Luyendijk and Margreet Steup-Beekman for their help with collecting the data. Furthermore we would like to thank Eveline Zijlstra for checking significant parts of the text for grammar and style.

REFERENCES

[1] ACR Ad Hoc Committee On Neuropsychiatric Lupus. The American College of Rheumatology Nomenclature and Case Definitions for Neuropsychiatric Lupus Syndromes. *Arthritis & Rheumatism*, 42(4):599–608, 1999.

[2] A. Artero, M. de Oliveira, and H. Levkowitz. Uncovering Clusters in Crowded Parallel Coordinates Visualizations. In *IEEE Symposium on Information Visualization*, pages 81–88. IEEE, 2004.

[3] S. Bachthaler and D. Weiskopf. Continuous scatterplots. *IEEE transactions on visualization and computer graphics*, 14(6):1428–1435, 2008.

[4] J. Blaas. *Visual Analysis of Multi-Field Data*. PhD thesis, Delft University of Technology, June 2010.

[5] J. Blaas, C. P. Botha, and F. H. Post. Interactive visualization of multi-field medical data using linked physical and feature-space views. In Museth;Ken, T. Möller, and A. Ynnerman, editors, *Proceedings of IEEE-VGTC Symposium on Visualization*. Eurographics Association, 2007.

[6] R. L. Brey, S. L. Holliday, A. R. Saklad, M. G. Navarrete, D. Hermosillo-Romo, C. L. Stallworth, C. R. Valdez, A. Escalante, I. del Rincón, G. Gronseth, C. B. Rhine, P. Padilla, and D. McGlasson. Neuropsychiatric syndromes in lupus: prevalence using standardized definitions. *Neurology*, 58(8):1214–1220, Apr. 2002.

[7] D. Cohen, S. P. Berger, G. M. Steup-Beekman, K. W. M. Bloemenkamp, and I. M. Bajema. Diagnosis and management of the antiphospholipid syndrome. *Bmj*, 340:1125–1132, May 2010.

[8] H. Doleisch, M. Gasser, and H. Hauser. Interactive Feature Specification for Focus + Context Visualization of Complex Simulation Data. *VISSYM '03: Proceedings of the symposium on Data visualization 2003*, pages 239–249, 2003.

[9] B. J. Emmer, T. W. J. Huizinga, and M. A. van Buchem. Neuro-Psychiatric Systemic Lupus Erythematosus. *MR imaging in white matter diseases of the brain and spinal cord*, pages 311–325, 2002.

[10] B. J. Emmer, S. C. A. Steens, G. M. Steup-Beekman, J. van Der Grond, F. Admiraal-Behloul, H. Olofsen, G. P. T. Bosma, W. J. N. Ouwendijk, T. W. J. Huizinga, and M. A. van Buchem. Detection of change in CNS involvement in neuropsychiatric SLE: a magnetization transfer study. *Journal of magnetic resonance imaging*, 24(4):812–816, Oct. 2006.

[11] B. J. Emmer, G. M. Steup-Beekman, S. C. A. Steens, T. W. J. Huizinga, M. A. van Buchem, and J. van Der Grond. Correlation of magnetization transfer ratio histogram parameters with neuropsychiatric systemic lupus erythematosus criteria and proton magnetic resonance spectroscopy. *Arthritis and rheumatism*, 58(5):1451–1457, May 2008.

[12] J. Fails, A. Karlson, L. Shahamat, and B. Shneiderman. A Visual Interface for Multivariate Temporal Data: Finding Patterns of Events across Multiple Histories. *2006 IEEE Symposium On Visual Analytics And Technology*, pages 167–174, Oct. 2006.

[13] Z. Fang, T. Möller, G. Hamarneh, and A. Celler. Visualization and exploration of time-varying medical image data sets. In *Proceedings of Graphics Interface 2007 on - GI '07*, page 281, New York, New York, USA, 2007. ACM Press.

[14] D. Gresh, B. Rogowitz, R. Winslow, D. Scollan, and C. Yung. WEAVE: a system for visually linking 3-D and statistical visualizations applied to cardiac simulation and measurement data. In *Proceedings of Visualization*, pages 489–492. IEEE, 2000.

[15] T. W. J. Huizinga and B. Diamond. Lupus and the central nervous system. *Lupus*, 17(5):376–9, Jan. 2008.

[16] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.

[17] D. F. Keefe, M. Ewert, W. Ribarsky, and R. Chang. Interactive coordinated multiple-view visualization of biomechanical motion data. *IEEE transactions on visualization and computer graphics*, 15(6):1383–1390, 2008.

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

[19] C. Plaisant, B. Milash, A. Rose, S. Widoff, B. Shneiderman, A. V. W. Bldg, and C. Park. LifeLines: Visualizing Personal Histories. In *Proceedings ACM CHI '96*, 1996.

[20] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23:S208–S219, Jan. 2004.

[21] S. C. A. Steens, G. P. T. H. Bosma, G. M. Steup-Beekman, S. le Cessie, T. W. J. Huizinga, and M. A. van Buchem. Association between microscopic brain damage as indicated by magnetization transfer imaging and anticardiolipin antibodies in neuropsychiatric lupus. *Arthritis research & therapy*, 8(2), Jan. 2006.

[22] E. N. G. M. Tan, A. S. Cohen, J. F. Fries, A. T. Masi, D. J. Mcshane, N. F. Rothfield, J. G. Schaller, N. Talal, and R. J. Winchester. Special article for the 1982 Revised Criteria for the Classification of Systemic Lupus Erythematosus. *Arthritis & Rheumatism*, pages 1271–1277, 1982.

[23] F. van der Heijden, R. Duin, D. de Ridder, and D. M. J. Tax. *Classification, Parameter Estimation and State Estimation: An Engineering Approach Using MATLAB*. Wiley, 1 edition, 2004.

[24] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman. Aligning temporal data by sentinel events. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 457, New York, New York, USA, 2008. ACM Press.

[25] T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee, and M. Smith. Temporal summaries: supporting temporal categorical searching, aggregation and comparison. In *IEEE transactions on visualization and computer graphics*, volume 15, pages 1049–1056, 2009.