UNIVERSITY OF CALGARY

PetroVis & FractVis: Interactive Visual Exploration of High-dimensional Oil and Gas Data

by

Ahmed E. Mostafa

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

July, 2013

Abstract

Many processes within the oil and gas domain deal with 'big data'–large sets of multidimensional data. Effectively analyzing these data sets is crucial to understanding the structure and the behaviour of oil and gas fields, and to the optimization of hydrocarbons production. However, experts face many challenges while attempting to analyze these data sets due to their high dimensionality, the inherent uncertainty of the data, and the lack of effective visual analytic interactive tools.

In this thesis, we attempt to look for new ways to support domain experts in interpreting high dimensional oil and gas data. For the exploration we designed, implemented and evaluated two new interactive visualization tools: FractVis and PetroVis. Our design efforts involved characterization of two oil and gas domain case studies, namely: microseismic monitoring (in the design of FractVis) and petrographic analysis (in the design of PetroVis). For each of the case studies, we outline the necessary tasks, needs, and the challenges faced by the domain experts. By closely collaborating with domain experts we iteratively designed, implemented, and evaluated the two interactive novel visualization systems to simplify the exploration of high dimensional domain data.

FractVis is a visualization tool aimed at supporting the visual analysis and exploration of the microseismic monitoring data. It combines, extends, and synchronizes parallel coordinates representation with other visualizations and interactions in order to facilitate the visual correlation of the data attributes. FractVis was further expanded by integrating new proxemic interaction and an interactive painting metaphor to simplify navigation and manipulation of the 3D microseismic data. The findings of our preliminary evaluation of FractVis suggest that the tool can provide insight regarding the simplification of the correlation of the microseismic data attributes, as detailed in the thesis.

PetroVis is a novel interactive visualization system developed for exploring petrographic data. PetroVis integrates interactive visualization elements with domain-specific statistical features, to

simplify the analysis process which involves (manual) validation of the automatic clustering of the data. The experts focus-group evaluation of PetroVis provided insight into the usefulness of the tool in simplifying the analysis of petrographic data clusters.

We conclude this thesis by presenting a set of design heuristics, reflecting on lessons we have learned while designing FractVis and PetroVis, with the hope of aiding and guiding future research that targets visual exploration of high dimensional oil and gas data.

Publications

Some of the materials, ideas and figures in this thesis have previously appeared in the following:

- **Mostafa, A.E., Greenberg, S., Brazil, E. A. V., Sharlin, E., Sousa, M.C. (2013) Interacting with Microseismic Visualizations, In Proc. ACM CHI 2013 Extended Abstracts - CHI'13. (Paris, France), 6 pages plus poster, April 27 - May 3. A CHI Work in Progress. Earlier version included within Report 2013-1035-02.**

- **Mostafa, A.E., Greenberg, S., Brazil, E. A. V., Sharlin, E., Sousa, M.C. (2013) Interacting with Microseismic Visualizations: The Video. In Proc. ACM CHI 2013 Video Program (Juried) - CHI'13. (Paris, France), April 27 - May 3.**

- **Cevolani, J.T., Mostafa, A.E., Brazil, E.A.V., de Oliveira, L.C., da Fonseca, L.G., Sousa, M.C. (2013) Computational methodology to study heterogeneities in petroleum reservoirs. In EAGE Annual Conference & Exhibition incorporating SPE Europec, SPE 164865-MS.**

- **Mostafa, A.E., Cevolani, J.T., Brazil, E.A.V., Sousa, M.C. (2013) Exploratory Visualization for Petrographic Characterization. Poster presentation, In the 4th Workshop on Interactive Data Visualization. June 6-7, 2013, FGV/EMAp - Rio de Janeiro, Brazil.**

- **Amorim, R., Mostafa, A.E., Brazil, E.V., Sousa, M.C., Eaton, D. (2012) Interactive Volume Reconstruction Based on Visualization of Multi-Attribute Microseismic Events, Microseismic Industry Consortium, Research Report plus Poster, Vol (1), Jan 2012, University of Calgary, AB, Canada.**

- **Mostafa, A.E., Amorim, R., Brazil, E.V., Eaton, D., Carpendale, S., Sharlin, E., Sousa, M.C. (2012) Exploratory Visual Modeling and Analysis of Microseismic Events, Extended abstract, GeoConvention / Vision, May 2012, Calgary, AB, Canada.**

- **Mostafa, A.E., Carpendale, S., Brazil, E.V., Eaton, D., Sharlin, E., Sousa, M.C. FractVis: Visualizing Microseismic Events. Paper accepted in the 9th International Symposium on Visual Computing. July 29-31, 2013 Rethymnon, Crete, Greece.** (To appear)

Acknowledgments

I have taken efforts in my masters. However, it would not have been possible without the kind support and help of many individuals. I would like to offer my sincere thanks to all of them.

I would like to express my special thanks of gratitude to my supervisor, Dr. Mario Costa Sousa, for all the guidance, support, and the constant supervision during my masters. I would like to extend my thanks to my co-supervisor, Dr. Ehud Sharlin, for his guidance, enthusiastic encouragement and honest advice. His willingness to give his time so generously has been very much appreciated. I would also like to acknowledge with much appreciation the input and support of the wonderful professor, Dr. Saul Greenberg.

A big thank you to my colleagues, Sowmya, Zamir, Jonatas, Felipe, Ronan, Elisa, Ali, Xiaoyang, and Nicole, for being supportive, for listening to me, and for providing me with wonderful suggestions. A special thanks to Juliana Cevolani for helping me in my research and for simplifying the oil and gas jargon for me. I would like also to thank Neda Boroumand for all the help regarding the microseismic monitoring evaluations. A special thanks and gratitude to Emilio who helped me during the whole study of my degree. Without your sincere suggestions, I would not make it. Thank you.

Last but not least, no words can describe the unprecedented support that I have received, and still receiving, from my family. I am sincerely grateful to my parents, my brothers, and my sister for the true encouragements.

It was an amazing learning experience and without all and everyone of you, I would not make it! Thank you :)

Table of Contents

List of Figures and Illustrations

Chapter 1

Introduction

The increasing global demand for energy has prompted the oil industry to encourage additional efforts to support the production optimization of hydrocarbons. As a result, the economical exploration and production of hydrocarbon fields are key strategic goals within the oil and gas domain. To achieve these goals, various sub-disciplines within the oil and gas domain perform operations wherein large amounts of data relating to oil and gas reservoirs are collected and analyzed. Making informed decisions about this data is a critical element for improving the quality and efficiency of reservoir explorations. In this regard, computational tools offer potential to assist and even dramatically improve the decision-making process and the exploration of reservoir data.

In most cases, analyzing the gathered data is a complex process involving the integration of various types of data sets with inherent uncertainty [90]. In addition, the scale of the gathered data is continuously growing, due to the use of better sensing devices and the accumulation of data collected over time, thus presenting high dimensional data spaces. Furthermore, the current exploration of reservoir data is often challenging since it also requires collaboration between various stakeholders who have different expertise levels [88]. Many believe that better computational tools can aid such experts when exploring and analyzing reservoir data. One important aspect in which automation can help is by providing better interactive visual analytic tools to aid with exploration of reservoir data. Within such visual analytic tools, 3D visualization, for instance, can support experts in exploring the 3D spatial geometry of the reservoir. Other visuals and interactions in these tools could also empower experts with different ways to correlate reservoir properties to better understand its behaviour.

Our research is focusing on the high dimensionality of oil and gas data with an objective of offering experts better interactive visualization techniques to facilitate its effective analysis. We employ a twofold approach; that is, attempting to (1) characterize the oil and gas data and the ways domain experts approach it; and (2) design interactive visual exploratory tools for analyzing this data.

1.1 Goal

The objective of this thesis is to design, implement, and evaluate interactive exploratory visualizations to simplify the interpretation of scientific high-dimensional oil and gas datasets. The approach we take in our investigation is multidisciplinary, and integrates in our design elements from the domains of microseismic monitoring, petrography, human-computer interaction (HCI), computer graphics, and information visualization.

The rest of the Introduction chapter is organized as follows: we begin by presenting, in Section 1.2, a brief overview of the two specific sub-domains we approached in our cases studies, namely microseismic monitoring and petrographic analysis. Then, in Section 1.3, we present the methodology used in our work. Section 1.4 outlines the contributions of this thesis and finally, an overview of the remainder of the thesis is presented in Section 1.5.

1.2 Background of the High-dimensional Oil and Gas Domains

The day-to-day activities of oil and gas engineers involve collecting and working with high-dimensional data. In their work, domain engineers often use commercial computational tools with traditional visualizations (such as scatter plots) in order to analyze and correlate their data. Yet, current (standard) visualization methods are not always sufficient in addressing the challenges (e.g., high dimensionality and uncertainty) experts face while analyzing the data. Therefore, our work integrates interactive visualization techniques in order to improve the experts' ability to explore oil and gas data. Our explorations are grounded in two case studies which provide specific oil and gas exploration context. The first case study concerns microseismic monitoring data analysis (introduced in Section 1.2.1), and the second case study focuses on petrographic analysis (introduced in Section 1.2.2). Both of these case studies share similar challenges and offer a unique potential for HCI and visualization research and are explored in more detail throughout the thesis.

1.2.1 Microseismic Monitoring Data Analysis

Microseismic monitoring is an important surveillance tool for reservoir development management [64]. Fracturing is at the core of microseismic monitoring and it is one way to create a reservoir allowing oil and gas trapped in rock pores to flow more easily [70]. Fractures are generated by injecting water or specially developed chemicals under high pressure inside the rock formation, causing it to break or fracture. Many seismic events result from this fracturing process and are called microseismic events. Detection of these events is done through an array of receiver systems (i.e. geophone sensors) distributed at specified locations near the injection point. However, due to limited acquisition geometry and geologic complexities and noise, uncertainty exists in this model [94]. Figure 1.1 shows an overview of the microseismic monitoring process, highlighting a single-well hydraulic fracturing and the microseismic events resulting from each of the multi-stage created fractures.

During the 1990's, microseismic monitoring has emerged as one of the most important tools to support decision-making within the oil industry. However, making informed decisions about improving an oil and gas reservoir model, based upon microseismic data, is a challenge for reservoir engineers and analysts. These difficulties arise due to the inherent features of the microseismic data, such as its intrinsic complexity, high-dimensionality, and a high degree of uncertainty. To help address these difficulties, microseismic experts are demanding better and more efficient interactive visualization tools that can adequately assist them in exploring the microseismic monitoring data.

Once microseismic monitoring data is gathered and processed, it is analyzed by domain experts such as geophysicists, geologists, and reservoir engineers. Each of these stakeholders represents a different skill set, and each often has different exploration goals, different interests, and various analysis tasks. Some important high-level tasks performed by these experts include understanding

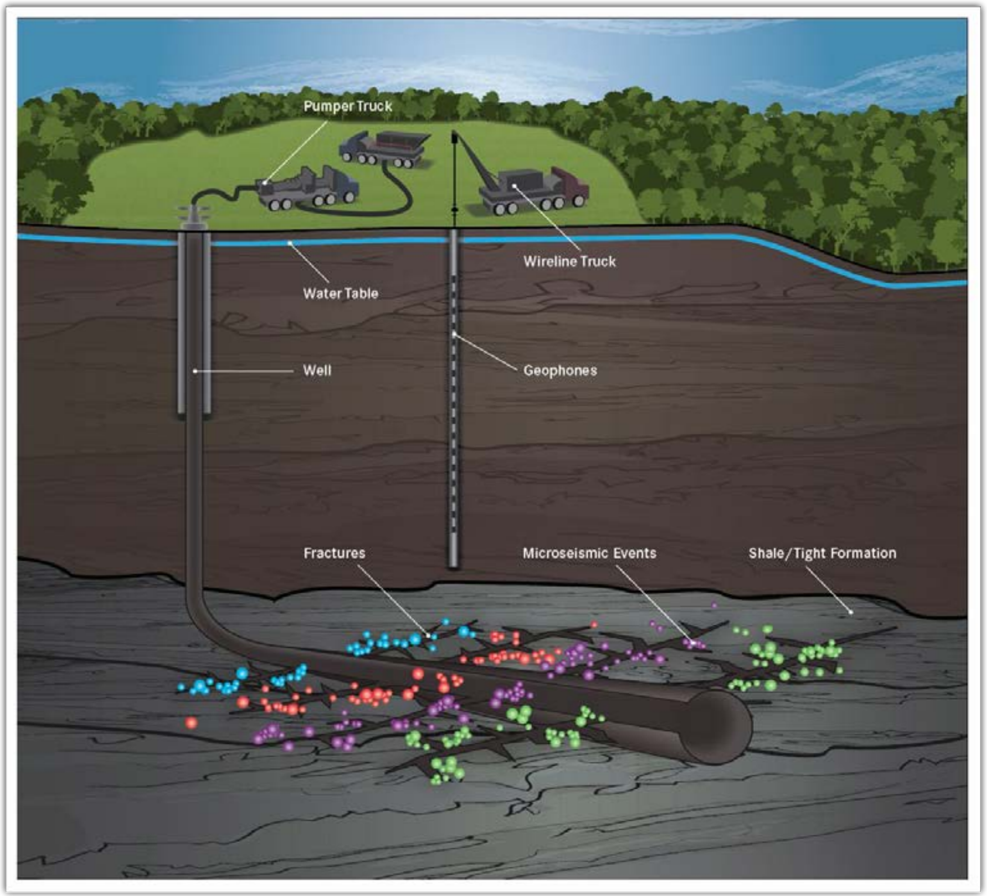

Figure 1.1: Hydraulic fracture schematic overview showing multi-stage fracturing (spheres in four different colors) along with a single well [2]

.

hydraulic fracture geometry, estimating stimulated reservoir volume (SRV), and optimizing long-term field development [70]. Most of these tasks have a 3D spatial component directly related to the exploration of microseismic (fracturing) geometry. For instance, expert interpreters need to know the 3D locations of the events in relation to the wells in the reservoir. This information could help experts filter out noisy events and perform correlations between various attributes within the massive dataset. Thus, these tasks could benefit dramatically from an interactive visualization exploration tool that converts the microseismic data into efficient and effective visual representations. Such a tool should be designed to better reflect and express the available information, the level of uncertainty, and other pertinent data details from different stages of oil/gas exploration and production. Indeed, we designed *FractVis* as a tool aiming to enable visual exploration of microseismic monitoring data (Chapter 3).

1.2.2 Petrographic Analysis

The characterization of hydrocarbon reservoirs is important to define the productivity of oil and gas fields. For a good characterization, it is necessary to understand the structure and the geometry of the reservoir. To help in this understanding, experts use many methods and techniques (e.g., petrographic analysis and seismic imaging) in order to analyze the rocks and learn more about their existing geometry. In particular, petrography is a sub-branch of geology that focuses on detailed description (and analysis) of rocks. However, analyzing such rock samples is a complicated and time-consuming task. Besides, proper studies demand integration and analysis of huge datasets, usually presenting high dimensional spaces. Currently, these difficulties are intensified due to the manual methods commonly used while analyzing the data. Consequently, we believe that automatic computational and interactive visual analytic tools can help experts in analyzing the petrographic data.

At the beginning of the petrographic analysis process, wellbores (or holes) are drilled to gather and collect rock samples at different depth (as shown in Figure 1.2). Then, laboratory slices (also called thin sections) of the collected rock samples are prepared for studying, and are examined using a petrographic microscope. The petrographic microscope is a type of optical microscope used by petrographic experts, to identify rocks and minerals within these thin sections. This process results in the identification of petrographic characteristics (also called petrofacies), which includes minerals and geological attributes organized in tabular compositional structure. Analyzing such petrofacies leads to a better prediction of any potential hydrocarbon storage in the reservoir [98].

Analysis and interpretation of high-dimensional petrographic data are usually performed using statistical methods and computational techniques, and domain experts are continuously seeking ways to improve these analysis methods, with the hope of improving the quality and outcome of the exploration. Currently, the petrographic computerized data analysis tools are not usually supported by intuitive interactions techniques, resulting in poor data exploration capabilities and often

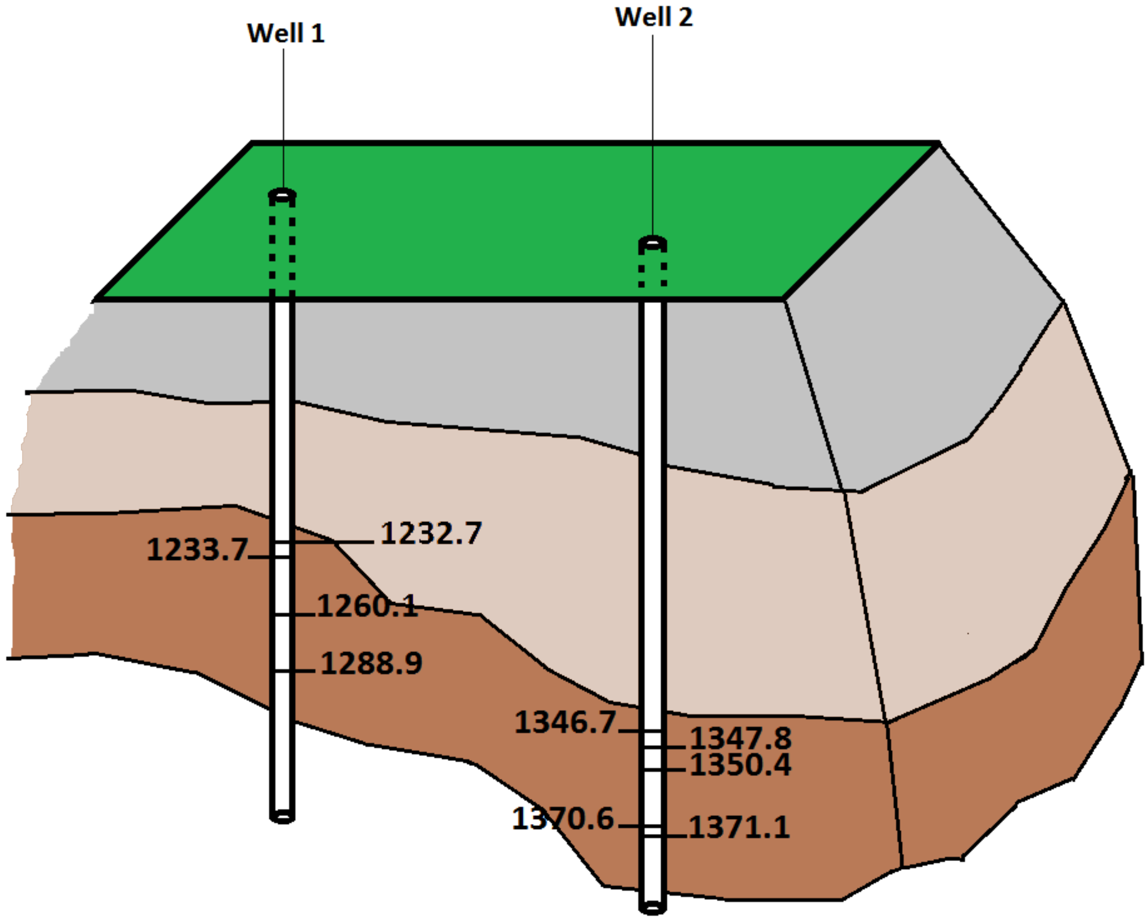

Figure 1.2: Schematic view showing wellbores and samples to be gathered at different depth (in meters). .

not properly utilize experts' knowledge and experience. For instance, experts need to manually integrate huge amounts of results to effectively interpret the data [24]. Therefore, we believe that interactive visual-analytic tools could dramatically benefit this process, providing experts with useful representations and effective tools as they explore and analyze their data. Indeed, we designed *PetroVis* as a tool aiming to enable visual exploration of petrographic data (Chapter 5).

1.3 Methodology

To address our research goals, we designed and implemented the following three visualization prototypes, iteratively revisiting our design based on continuous feedback from domain experts:

- *FractVis*: an interactive visualization system developed to enable exploration of microseismic data by adapting and extending scientific and information visualization techniques. (*FractVis* is discussed in details in Chapter 3)

- *Proxemic FractVis*: a prototype based on *FractVis* using proxemics and 3D interaction techniques, tailored to facilitate exploration and manipulation of 3D microseismic data.

- *PetroVis*: an interactive visualization system developed to support petrographic experts while analyzing their data. (*PetroVis* is discussed in detail in Chapter 5)

Prior to the development of our prototypes, our first task was to learn more about the high dimensional oil and gas sub-domains of microseismic monitoring and petrographic analysis. To achieve this, we followed an iterative design approach involving characterization and requirement analysis of the aforementioned domains. We conducted continuous meetings and consultation sessions with domain expert collaborators in order to learn more about the domains, the related terminology and jargon, and to understand the experts' tasks, processes, challenges and needs. Several of these meetings take inspiration from observational techniques such as rapid ethnography [66] and contextual inquiry [77]. Our characterization of the domains (detailed in Chapters 3, 4, and 5) includes description of the domain data and analysis of the experts' common tasks.

Following the domains' characterization, we designed and developed *FractVis* and *PetroVis* by applying interactive visualization techniques to the high-dimensional domain data. We integrated several novel elements and interaction techniques, such as parallel coordinates (PCs), into our visualization applications to better support experts in analyzing their data. For example, the visualization of PCs within *FractVis* used color-based correlation to easily compare data attributes without the need to perform axis reordering. *Proxemic FractVis* (our third prototype) is using proxemics and 3D interaction techniques to facilitate the exploration of 3D microseismic data. By applying these interaction techniques, we attempted to help microseismic experts to more easily navigate and manipulate their data based on the insight we had gained regarding their goals, tasks and workflow.

FractVis and *PetroVis* were evaluated by domain experts through a set of design critique and focus group sessions. In addition, visualization researchers (computer science experts) have participated in the evaluation of *FractVis* to assess its visualization strength and weaknesses. The outcome of these evaluations highlighted various insights helping us in understanding the usefulness of the developed prototypes. For the *Proxemic FractVis*' prototype, the evaluation involved a set of design critique sessions that enabled us to reflect on the design challenges and the potential benefits of adapting such interaction techniques in the future.

1.4 Contributions

The main contributions of this thesis are:

1. **Design, prototyping, implementation and evaluation of *FractVis* - a novel interactive visual analysis and exploration tool for high dimensional microseismic monitoring data**.

2. **Design and implementation of *Proxemic FractVis* – a novel interactive prototype exploring the application of proxemic interaction and 3D interaction techniques in the domain of microseismic monitoring data analysis**.

3. **Design, prototyping, implementation and thorough evaluation of *PetroVis* - a novel interactive visual analysis and exploration tool for high dimensional petrographic data**.

4. **A set of design heuristics for future design efforts in the domain of interactive visualizations of high-dimensional oil and gas data**.

1.5 Thesis Overview

The remaining of this thesis is organized as follows:

- In *Chapter Two*, we provide an overview of the key related work regarding different elements of this thesis. We discuss a number of relevant recent efforts in the realm of visualization, computer graphics and HCI. We also briefly present an overview of the oil and gas domain with particular emphasis on the high dimensional processes of microseismic monitoring and petrography.

- In *Chapter Three*, we present the design, implementation, and evaluation of *FractVis*: a visual exploratory prototype, developed to enable visual analysis and exploration of microseismic data.

- In *Chapter Four*, we introduce *Proxemic FractVis*, a tool exploring new forms of interaction with microseismic data, allowing experts with more natural techniques of exploring the spatiality and the fracture geometry of the 3D microseismic data.

- In *Chapter Five*, we present the design, implementation, and evaluation of *PetroVis*: an interactive visual analytic prototype, developed to support the visual analysis of petrographic data.

- In *Chapter Six*, we present a set of design heuristics that emerged from our research, in a hope to guide future efforts of designing interactive visual explorations of high dimensional oil and gas data.

- In *Chapter Seven*, we conclude the thesis, and highlight perspectives for future work.

Chapter 2

Background

The theme of this thesis contains techniques and concepts borrowed from information visualization and human computer interaction. Using such concepts, we present an exploration of microseismic monitoring and petrographic data sets, as two case studies within the context of oil and gas domain. Each of the case studies covers a detailed description of the visualization and interaction techniques adapted to simplify the data analysis and exploration. The following subsections review existing literature in order to establish the necessary background regarding different components of our work. This chapter is organized as follows:

1. A discussion of key related works regarding different aspects of building visual-analytic systems including characterization of domain problems, and integration of visualization techniques.

2. A brief background regarding 3D interaction techniques combined with an overview showing the potential of how recent concepts such as proxemic interaction can be useful to simplify interaction within 3D spaces.

3. A short introduction to oil and gas exploration and production cycle with emphasis on the examples of microseismic and petrography as two high-dimensional domain instances.

2.1 Visual Analytics

The field of visual analytics is a growing field with big potential for simplifying the analysis of continuously growing datasets [51]. The goal of visual analytic systems is to support decision-making and enable users to obtain deep insights about their data, by adapting effective forms of visualization and interaction. In general, building successful visual analytic systems begins with

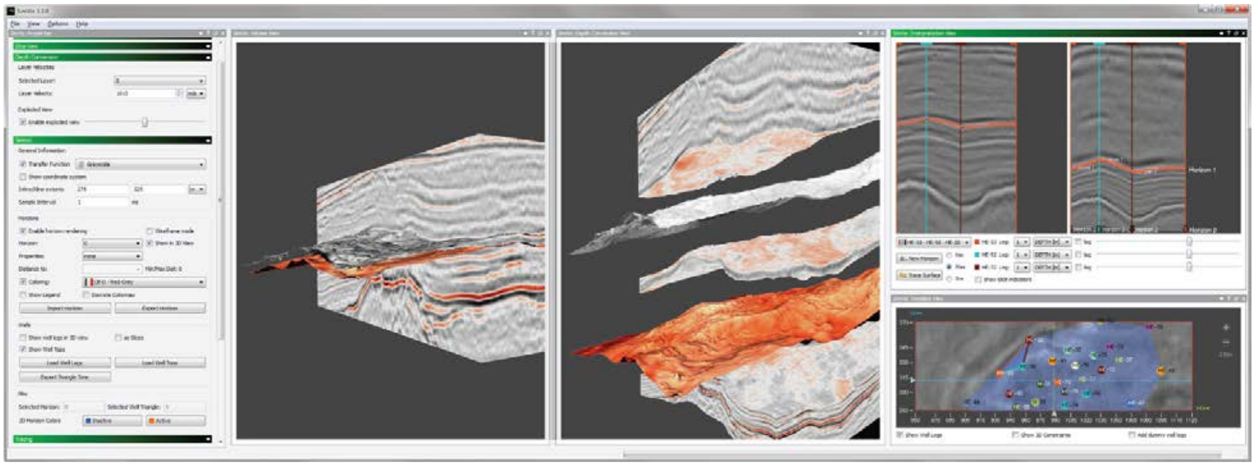

Figure 2.1: Interactive visual analysis of seismic horizons surfaces in reservoir volumes [42], showing a 3D view in the middle representing the volume alongside well positions, and a complete horizon surface, and other visuals for showing horizon surface slices .

characterization of domain problems. After that, implications could be outlined to guide the design and development of the actual systems. This includes leveraging effective forms of visualization and interaction in order to empower users and simplify their data exploration. This process should be iterative and involve continuous collaboration with domain experts. Finally, a proper evaluation of the developed system is important to spot any found insights [83]. Recently, many systems and technologies have been developed for the purpose of data exploration [42, 45, 75, 80, 33]. For instance, in Figure 2.1, we an example of an interactive visual analytic system to explore seismic horizon surfaces in reservoir volumes. Similarly, in figure 2.2, we see a visualization aiming to support exploration of field measured seismic data.

Our work aims at assisting experts in making informed decisions, as an effort to address the lack of computational and visual analysis tools of geophysical and geological data in the domain instances of microseismic monitoring and petrography. Following we review key related works involved in the process of building visual-analytic systems. We also explain how they are related to the context of the chosen domain instances.

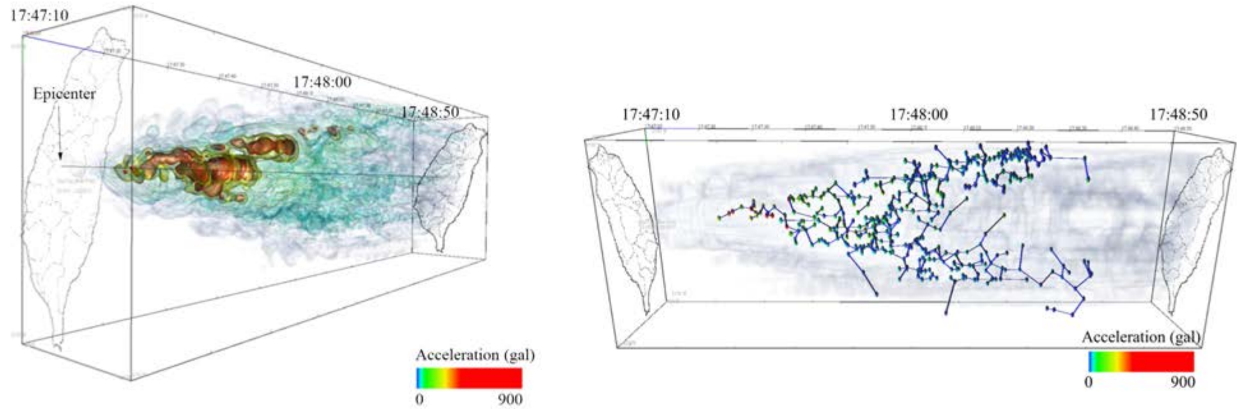

Figure 2.2: Visualizing field-measured seismic data [45], showing ground-motion volume data through volume rendering.

2.1.1 Domain Characterization

Using field research methods before and during the actual development of any system is very useful to characterize the system domain and understand the context of the environment [83]. Ethnography, for instance, is a well-known method to learn about any domain, its practices, needs and problems in order to discover if and how visualization can enable insight and discovery. Such learning can be acquired through different techniques including talking to (interviewing) and observing domain experts. However, the use of ethnographic techniques usually takes a long-time to properly understand the domain and its users' activities. Rapid Ethnography [66] is a recent technique proposing different strategies to accelerate the domain characterization process. Similarly, Contextual Inquiry [77] is another field method used to gather information about domain users and their specific needs.

The essence of such methods allows for gathering user requirements, developing user models, abstracting the data and identifying any challenges. Many successful systems used these methods. For instance, Meyer et al. [65] described a problem in the field of comparative genomics and developed a visualization to simplify genomes data exploration. As we can see in figure 2.3, MizBee is a visualization browser showing different synchronized visuals of the genomes and their associated data. Successful examples from other domains proposed a complete visual-analytic

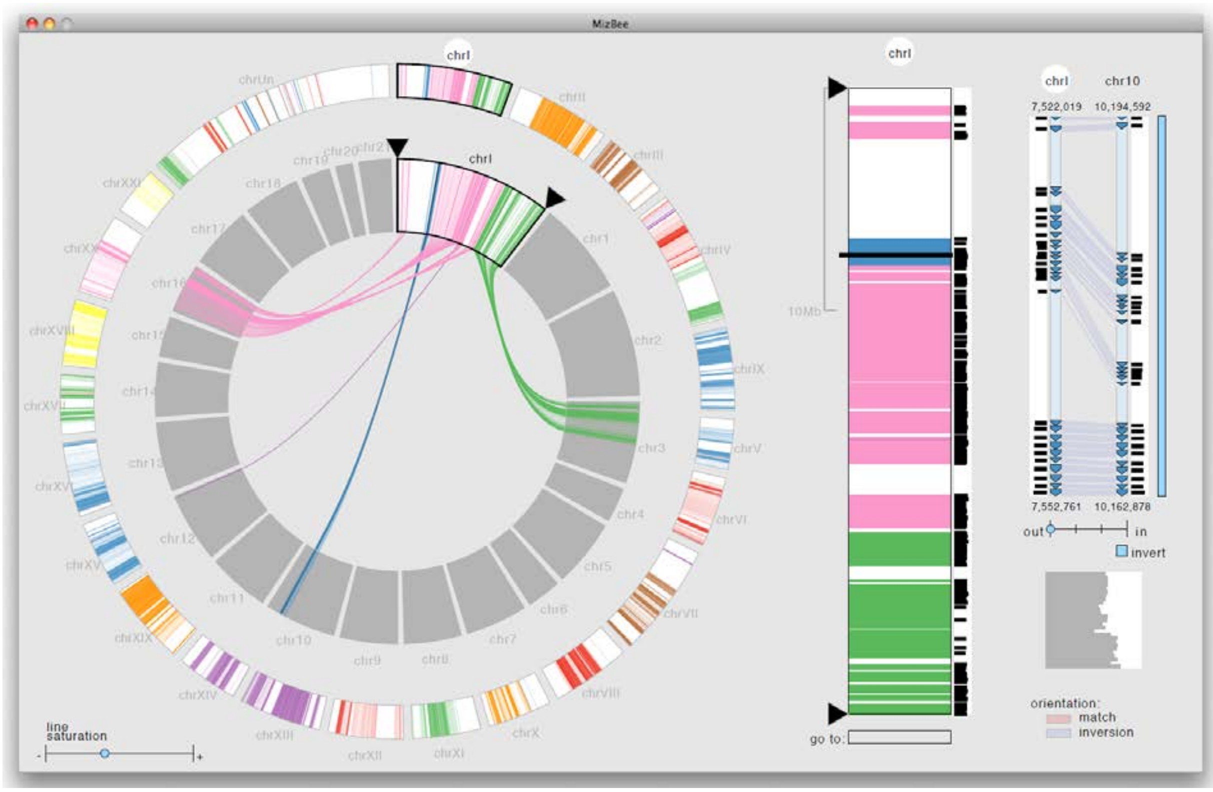

Figure 2.3: Exploratory visualization of comparative genomic data [65], showing the multiscale MizBee browser with different linked views at of the genome, the chromosome and block levels.

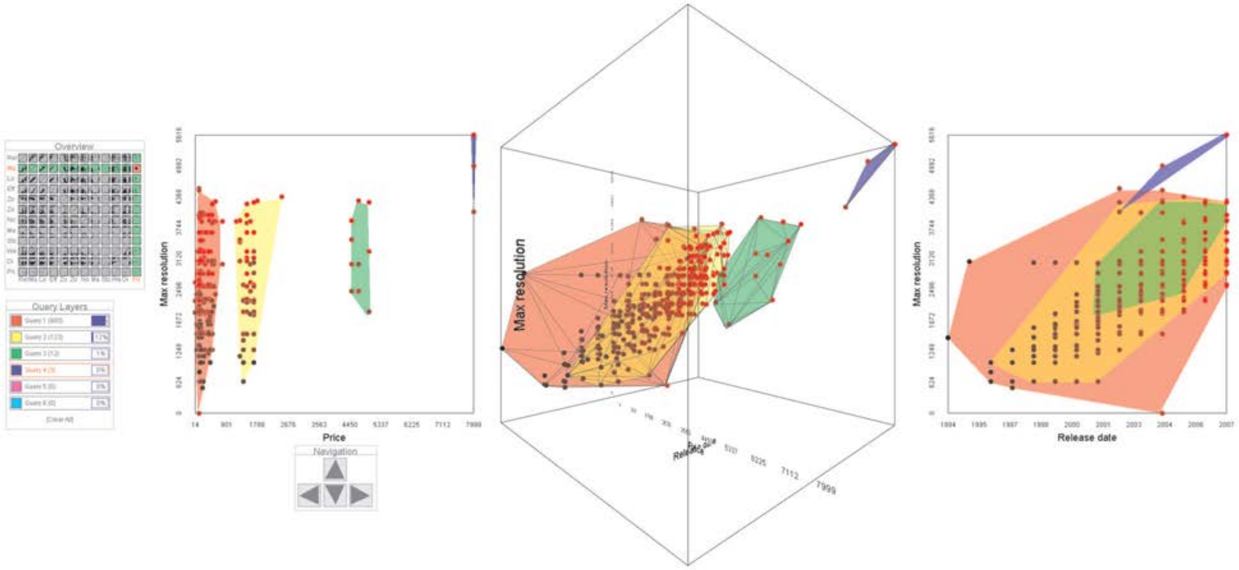

Figure 2.4: Multidimensional visual exploration of a digital camera dataset using scatterplot matrix [36], supported by an animated 3D rotation the visualization highlights the transition while the user is querying the data.

system, including a characterization of the domain as well as a visualization prototype [68, 82]. On the other hand, other systems focused on problem characterization as their main contribution [56, 93], and they did not provide a visualization-based solution to simplify the problem domain. In our work, we are inspired by the techniques of contextual inquiry and rapid ethnography while learning about the two oil and gas sub-domains: microseismic monitoring and petrography.

2.1.2 High Dimensional Exploration

Multidimensional visualization involves working with high dimensional data sets. The structure of such datasets usually contains many variables (or dimensions) for each data sample (or observation). Visual exploration of these data sets is important in many scientific applications that demand finding hidden relationships between the data attributes. In general, such exploration may take different forms involving finding data trends, highlighting data patterns, detecting outliers, or correlating the data attributes. In this section, we review techniques and methods developed in information visualization literature for visualizing multidimensional data.

Most of the high-dimensional techniques can be classified according to how they manipulate the dimensionality of the data. They could be projective, non-projective, or dimensionality-reduction techniques [100]. Furthermore, some of these techniques are more suitable than others when exploring certain types of data, and it is best determined according to the dataset's scheme and the intended usage scenario. For instance, a visualization researcher may suggest the use of projective techniques for analyzing the high-dimensional data if the aim of the exploration is to focus on the neighborhood relationships among the data points. Our approach adapted mainly a non-projective visualization technique because we believe it is more suitable for the representation of the studied domain-data. Following, we summarize some of the multidimensional visualization techniques and the key related works that inspired our implementations.

A scatterplot is a simple projective technique which graphically plots two variables of the high dimensional data by mapping them to the cartesian coordinates [28]. A scatterplot matrix [36] is a multidimensional data visualization technique which creates a matrix of N^2 scatterplots arranged in N rows and N columns. Figure 2.4 shows an example of visual exploration of a multidimensional digital-camera dataset using the different interaction integrated through the use of scatterplot matrix. However, the resolution of each scatterplot in the scatterplot matrix is limited when data contains very high dimensionality. Although this problem can be simplified using the automatic sorting of the coordinates [76, 101], it still requires a profound knowledge of the multiple coordinated views and their synchronizations.

Dimensionality reduction methods, on the other hand, focus on mapping high-dimensional spaces to low-dimensional spaces, and attempt to preserve (distance) relationships between pairs of points in the high-dimensional space. Principle Component Analysis (PCA) [48] is one of the well-known techniques that can be used, generally, to describe the dominant trends in the data through a number of orthogonal dimensions. Multi-dimensional scaling (MDS) [99] is another popular method that relies on relational measures between pairs of data samples. This technique, however, focuses mainly on studying similarity of data samples and tends to be statistically in nature, which may not be the best approach for simplifying the visual analysis of all types of data.

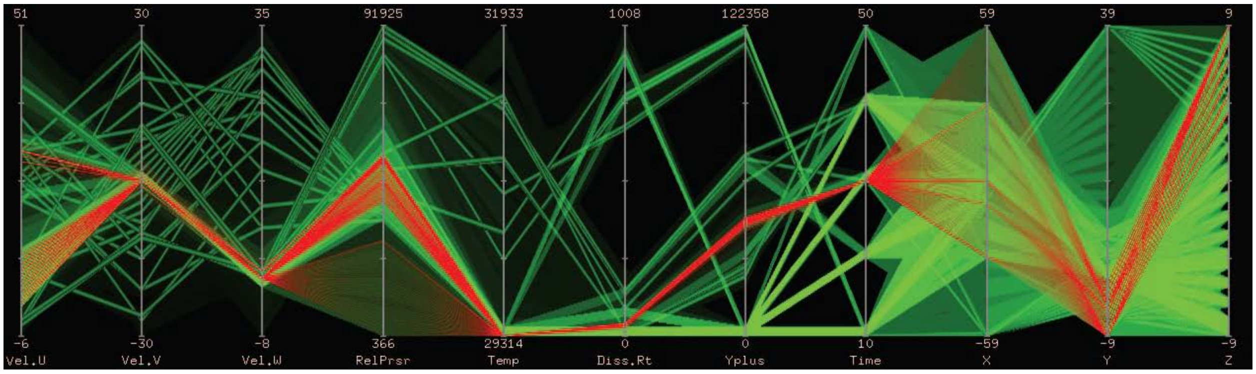

Figure 2.5: Outlier-Preserving Focus+Context Visualization in Parallel Coordinates [71], highlighting the outliers that are very visible along with major trends of a flow simulation dataset.

Axis reconfiguration (non-projective) techniques are another type of techniques which map each data point to a glyph. Star Coordinates is one variation of axis-based visualization in which all data dimensions share the same origin [49]. Similarly Star plot (or Radar Chart) [25] represents each observation as a star-shaped figure by showing one ray for each variable. Elmqvist et al. [37] proposed a star plot-based system with a visual canvas to support the analysis of large-scale multivariate data with flexible visual queries. Parallel coordinates (PCs) [47, 46] is another axis-based well-known visualization technique for representing highly dimensional data in a 2D plane. The technique of PCs represents each dimension of the data as vertical parallel axis, and each data samples as a polyline intersecting each vertical dimension proportionally to the sample's value.

PCs and many axis-based techniques may suffer from the problem of visual cluttering. However some strategies exist, such as brushing [62] and axis reordering [55, 76] to alleviate and simplify these problems. In addition, focus-plus-context techniques [23, 38] can be integrated with these strategies to highlight certain features [71] in the data and reduce the cluttering. For example, in the figure 2.5, we see a visualization of the parallel coordinates with outliers being highlighted in different colors to facilitate outliers' identification. Visualization systems that contain Focus-plus-context visuals enable users to see the information of interest in the foreground and all the remaining information in the background at the same time, in a single display. The figure in 2.6 highlights a visualization of a map using the technique of focus+context wherein the

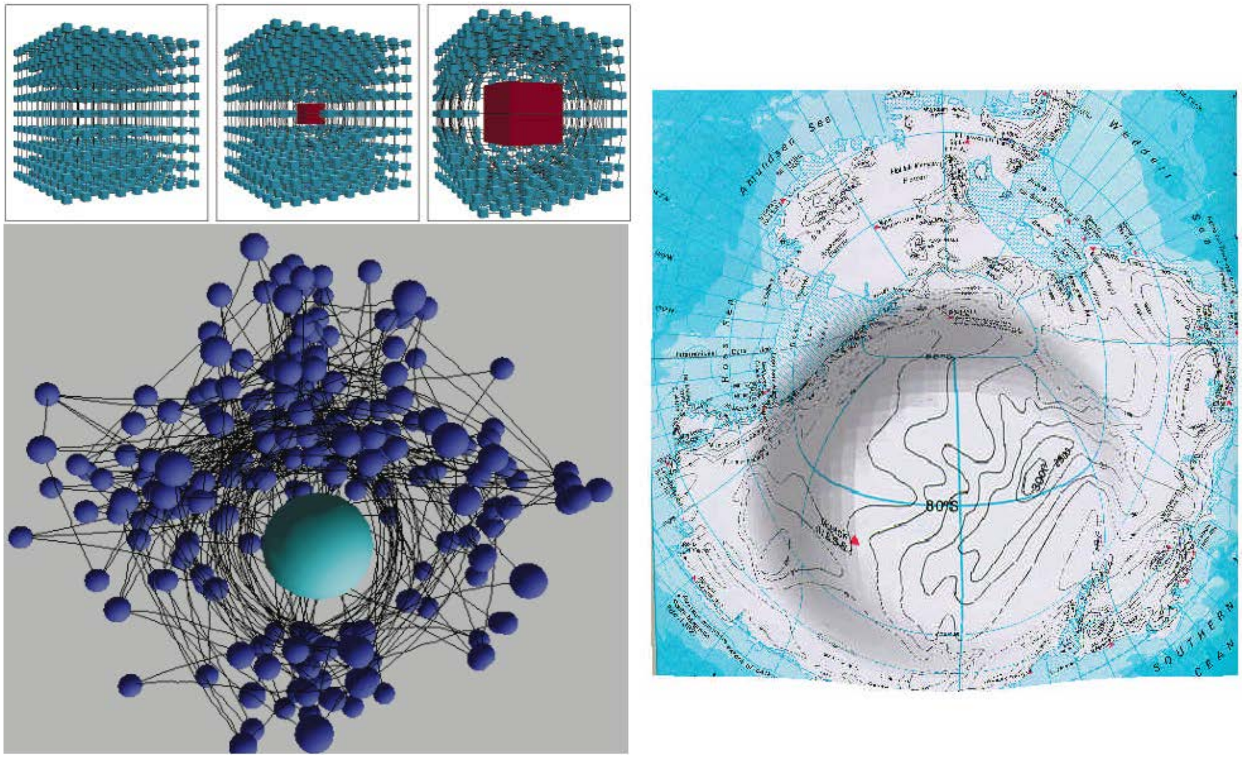

Figure 2.6: Examples of Focus and Context: (left) Extending distortion viewing from 2D to 3D [22], and (right) presentation of focus and context applied to map view [23].

part in focus is has extruded up to reflect zooming and presenting more details while considering the remaining of the map as context.

Different visualization systems have used the technique of PCs for data exploration. XmdvTool [97] is a general system that supports high-dimensional exploration by combining different multivariate visualization methods including the technique of PCs. Steed et al. [87] presented a system for analyzing weather data using an enhanced PCs' implementation (Figure 2.7). Martin et al. [62] discussed high dimensional brushing for exploring multivariate data with focus on PCs. Siirtola and Kari-Jouko [86] also evaluated PCs, and the results highlighted that the users found PCs more effective than other methods when they performed their tasks.

Integrating PCs with visualization techniques and interfaces has been explored to support and enable the development of better visualization tools. For example, Zhang et al. [104] integrated PCs with custom visuals (e.g., networking-based interface) to explore and analyze high dimen-

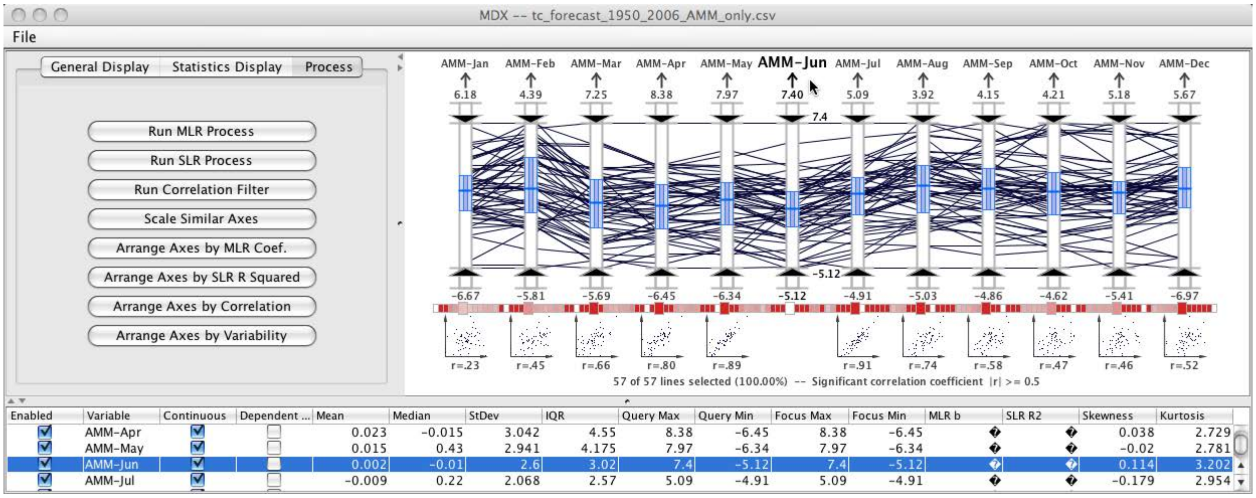

Figure 2.7: Visual analysis of weather data using enhanced parallel coordinates implementation [87], showing the possible interactions while selecting some data attribute while intending to perform correlation with the other attributes.

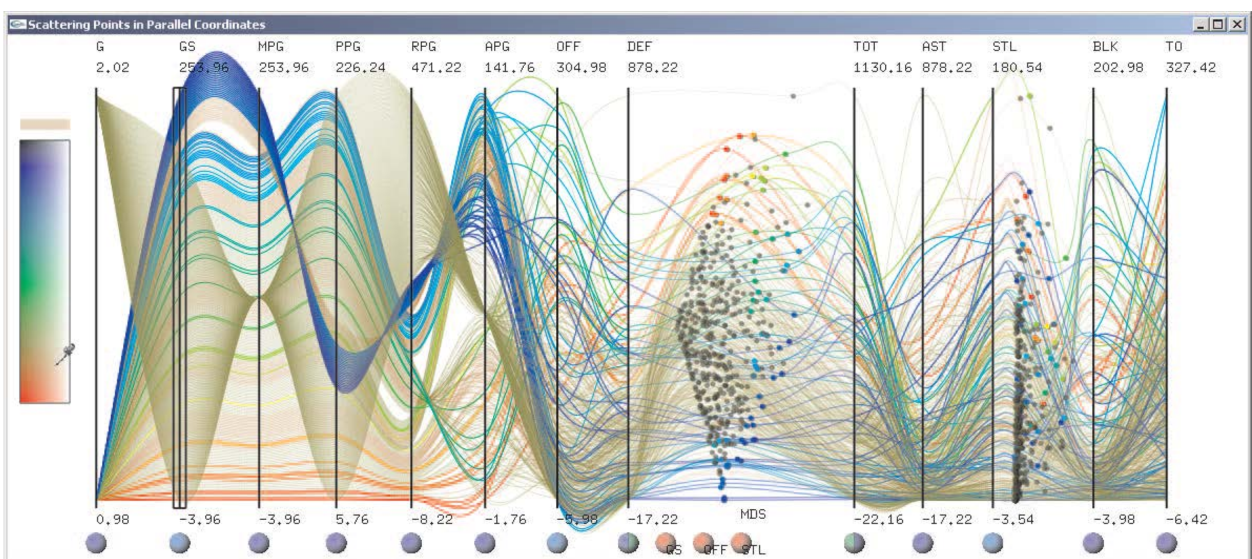

Figure 2.8: Scattering data points within parallel coordinates visualization [102].

sional networking data. Yuan et al. [102] presented a system that scatters the data points using MDS within the visualization of PCs (Figure 2.8) though a GPU-based implementation to facilitate data navigation and interaction. In addition, Holten and van Wijk [43] conducted a user study to evaluate PCs when applied to cluster identification, and they integrated PCs with dynamic boxes and scatter plots, among other techniques. The findings of their study suggested that most of PCs improvements, with the exception of embedding scatter plots within PCs, do not result in significant performance gains. We extended this idea by supporting interactive integration of dynamic magic lenses [16] within the PCs' visualization to provide flexible way of interaction; for instance, to filter the data in a more intuitive way. Indeed, we adapted many of the aforementioned ideas, and extended the PCs to simplify the exploration and analysis of the domain data, through the prototypes that we developed.

2.1.3 Multiple Coordinated Views

Multiple coordinated views is a known technique in many visualization systems that require synchronization and presentation of data in different ways simultaneously. Bowman et al. [18], for instance, presented an example of a system for analyzing MRI repositories using multiple coordinated views. Roberts [78] provided a discussion of the state of art on using multiple coordinated views, and he included different example systems that incorporated this technique. Furthermore, a set of guidelines for using multiple views in information visualization has been proposed by Wang-Baldonado et al. [95]. On the other hand, Andrienko et al. [13] provided a critical examination of multiple coordinated views. Similarly to systems that support coordination of different representations of data, we also used multiple coordinated views in our prototypes (detailed in Chapters 3 and 5). Figure 2.9, for instance, shows an overview about such coordination in the early prototype of FractVis (Chapter 3).

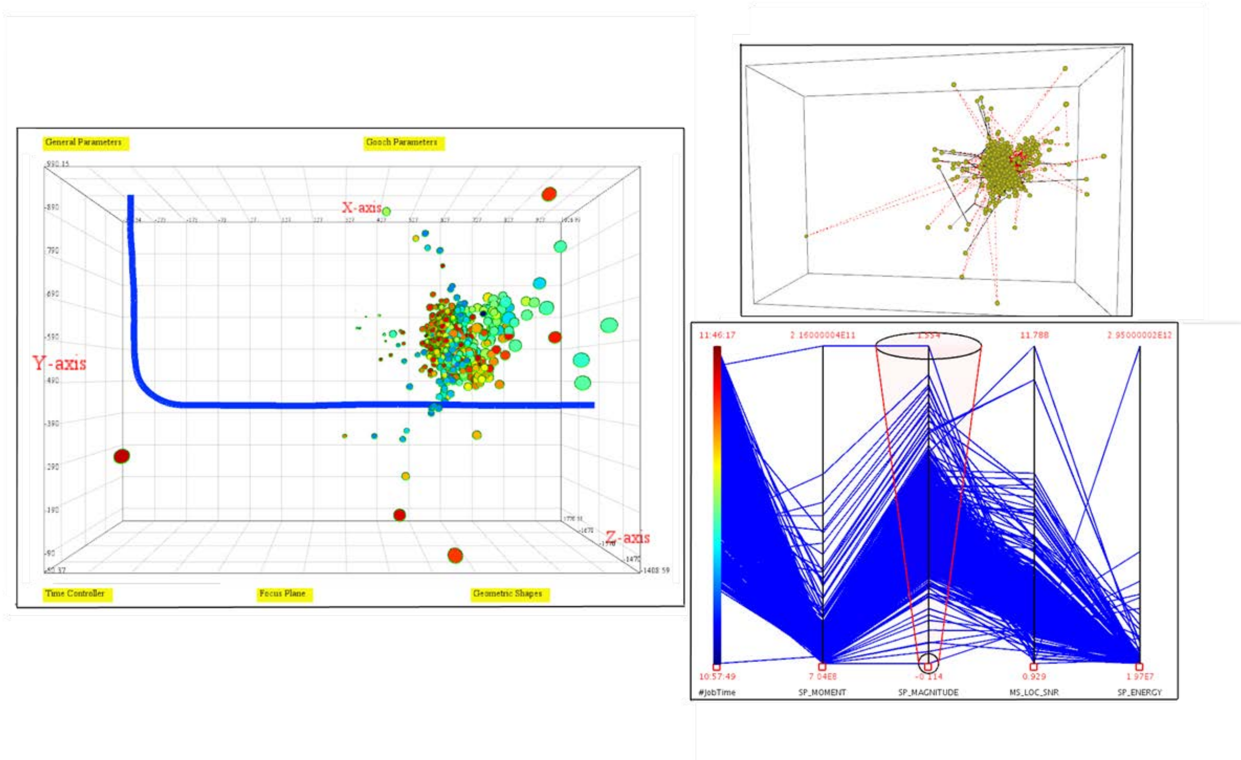

Figure 2.9: Visualization of microseismic events involving 2D and 3D coordinated views [67].

2.2 Interaction for 3D Exploration

It is essential to simplify the user interface of visualization systems in order to lower the learning curve for users. Our microseismic visualization prototype (Chapter 3) uses 3D data and we want to simplify navigation and interaction with such content. Therefore, we investigated recent interaction techniques with the goal of simplifying exploration and manipulation of the 3D data. In this section, we review 3D interaction techniques. Additional description of 3D techniques can be found in chapter 4.

2.2.1 Proxemics and 3D Interaction

Nowadays we can visualize and render virtual content with reasonably high quality. The value of using and interacting with such visualization will be easier and more effective if it leverages our natural abilities similarly to real world interactions. However, interacting with 3D content in a

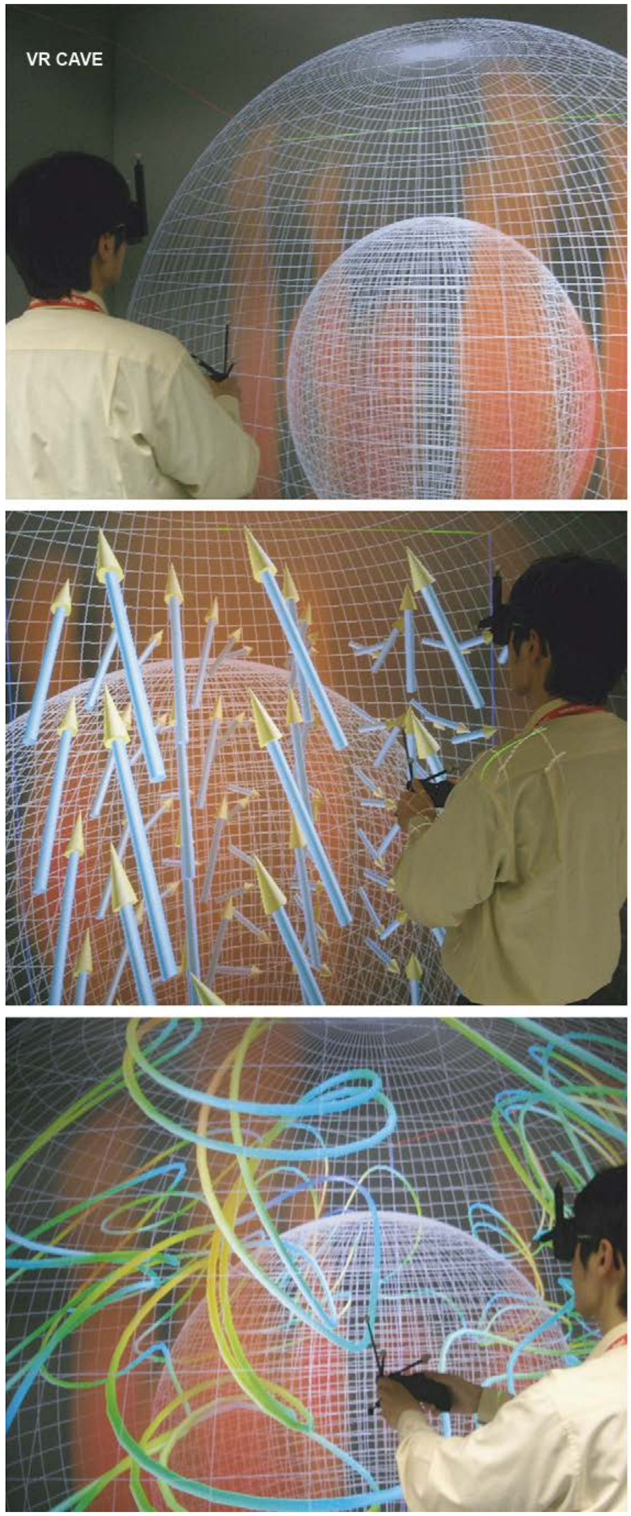

Figure 2.10: Scientific visualization of geophysical simulation data by the CAVE VR system with volume rendering [73].

Figure 2.11: Fish Tank VR for scientific visualization application [52]

natural way is still problematic. Such issues inspired many researchers to develop and investigate various techniques for improving 3D interactions. Various techniques for spatial navigation have been extensively researched, where their goal is to allow users to access and manipulate 3D entities using techniques that sometimes borrow (interactions) from the physical world. A survey about 3D interaction techniques and their history is proposed by Hand [41] with considerable focus on navigation and object manipulation in 3D virtual environments. Following is a brief description of the techniques that inspired our work including virtual reality (VR) [19] and proxemic interactions [15].

VR [19] is one of the classic 3D interaction techniques that reflect more realistic interactions. In one form of VR techniques, people manipulate avatars of themselves that simulate one's physical presence within a completely virtual and synthetic environment. Caves [30] are other VRs which use projective technologies to surround and immerse a person within the 3D space. For example, figure 2.10 shows a VR-based visualization of a scientific simulation of geophysical data. In such systems, 3D data is seen either on large multiple displays or Stereo glasses. Directional sound and input devices, such as data gloves, can be adapted to enrich the 3D experience even further [81]. Fish Tank [14] is another VR technique which immerse the user by redrawing the 3D stereo scene

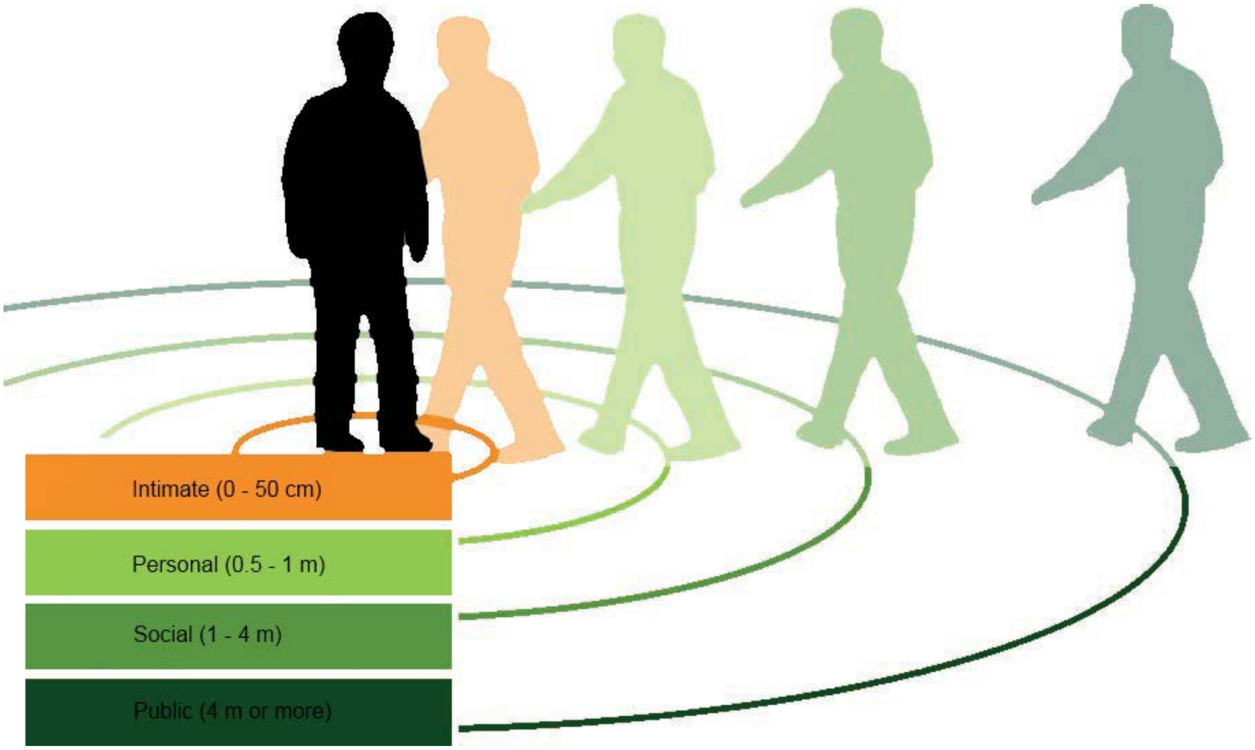

Figure 2.12: Edward Hall's proxemic zones. Hall correlates physical distance to social distance between people and categorizes it into four discrete zones.

on the screen depending on his head's position in space. Unlike CAVE, Fish Tank VR creates an illusion of a "real" third dimension. Such technique, however, is a single-user technique because non-head-coupled users will not see correctly the 3D image since it is not computed for them. Some researchers evaluated CAVE and Fish Tank VR and provided a subjective comparison about the pros and cons of each of them [52] (Figure 2.11).

Recently within HCI, the use of spatial relationships has emerged as another interaction modality to improve communication and interaction with everything in our everyday environments. Edward Hall described proxemics as a theory about correlating interpersonal spatial relationships [40]. He talked about correlating physical and social distances and the role of proxemics as a form of interaction that is natural and understood by people but yet to be fully realized by interactive computing systems. Few systems incorporated a simple use of proxemics, for instance, to detect presence or absence of people and/or devices in the environment [96]. Other recent systems such as Vogel [92] applied directly Hall's social distances (four zones as in Figure 2.12) to

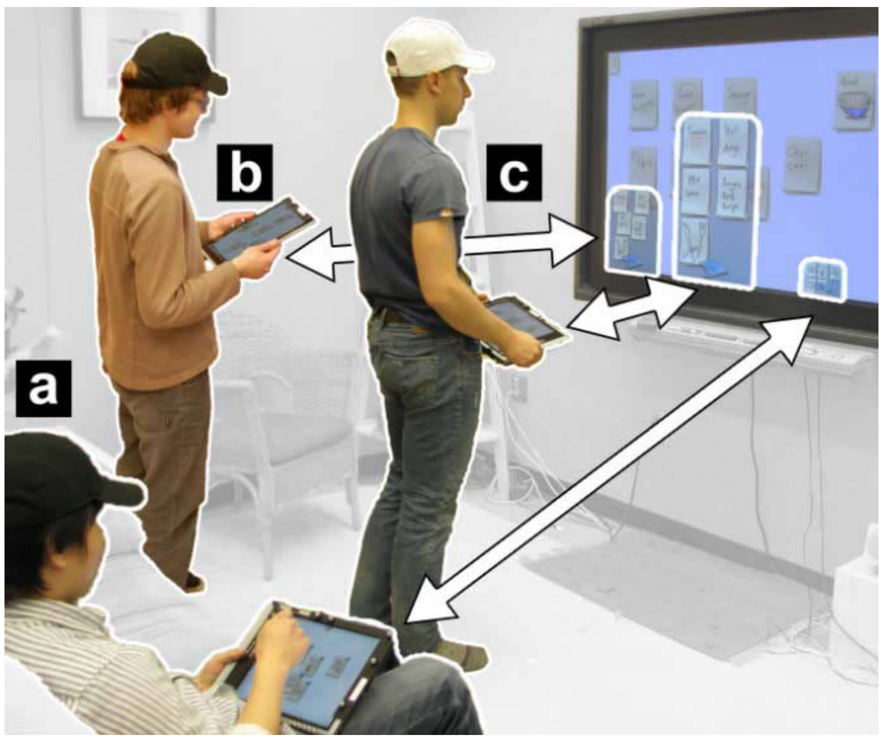

Figure 2.13: Proxemic Interaction to facilitate information exchange between digital devices as a function of proximity, showing how information can be progressively revealed and transferred following awareness and gradual system reaction [59].

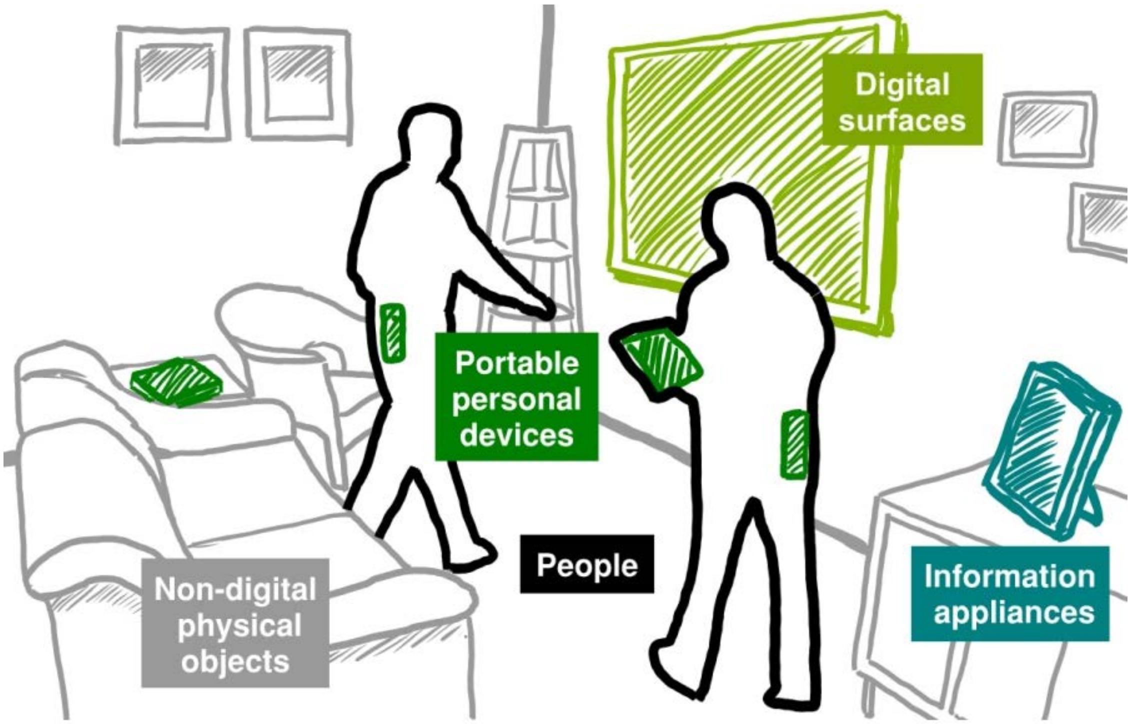

Figure 2.14: Proxemic interactions in ubiquitous computing ecologies, wherein relations are considered between people to devices, devices to devices, and non-digital physical objects to people and devices [58].

extend digital interfaces to react to people and update the information displayed. Marquardt and Greenberg [59, 61] extended this concept further and introduced *proxemic interaction* by using people's natural expectation of distance to mediate interaction. As we can see in figure 2.13, system awareness of people and devices with gradual engagement is illustrated. In essence, the notion of proxemic interactions relates people to devices, devices to devices, and non-digital physical objects to people and devices as we can see in figure 2.14. In our work, We adapted 3D interaction techniques involving the use of Wiimote to paint and manipulate the data, similarly to the work of [54] and [53] (in Chapter 4). We combined such techniques with the proxemic interactions with the goal of simplifying interaction with the 3D microseismic data.

2.3 Oil and Gas Exploration Overview

Oil and gas reservoirs are subsurface volumes representing potential storage of hydrocarbons contained in porous or fractured rock formations. The goal of studying hydrocarbon reservoirs is the exploration and optimal extraction of hydrocarbons in the most economical and environmental friendly manner. Exploration and production cycle of hydrocarbons involves many complex processes and operations that happen (usually) during the life of the well. Many of these processes are inter-disciplinary due to the interdependence of their data sources. While it is important to organize and analyze these huge amounts of data, it is very challenging due to the increased complexity and information stored in the raw data. The focus of this research is on two different sub-domains: microseismic monitoring and petrographic analysis. These two examples involve the use of multidimensional data and they share similar challenges regarding data interpretation and analysis.

Microseismic monitoring has been used for decades for various applications including engineering, mineral mining, and water storage. The technology has existed in the Petroleum industry for decades, but more recently has gained much interest. The importance of this new method is increasing due to the focus of the industry on improving reservoir production [64].

Petrographic data represent detailed description of rocks. Analyzing such data supports characterization of the hydrocarbon reservoir leading to better exploration and production (E&P) of oil and gas fields. However, reservoir characterization involves many challenges due to the high dimensionality of the data as well as the manual process used in the data analysis [24].

We performed characterization of both domains. Then, we designed and developed visualization systems (in Chapters 3 and 5) tailored to the requirements of their experts. We developed *FractVis*, a visualization system to simplify the selection of microseismic events in order to refine the analysis and estimate the stimulated reservoir volume (SRV). We also developed *PetroVis*, a visualization system for petrographic data analysis in order to simplify petrofacies' (clusters) validation and qualification.

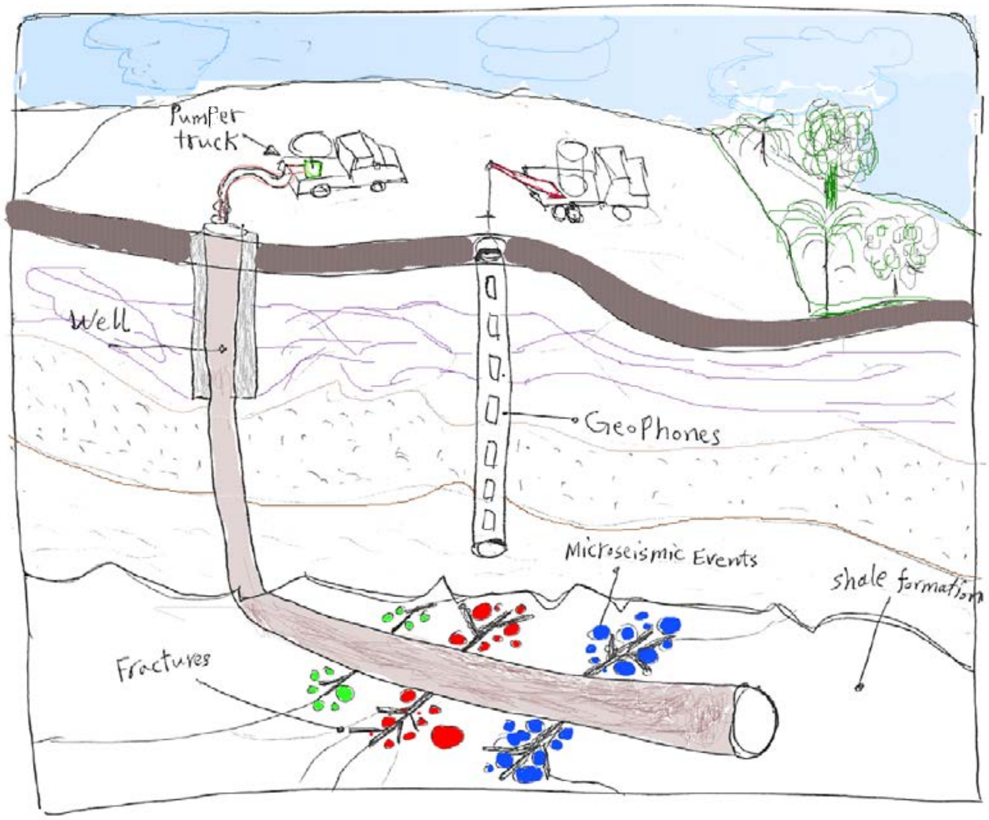

Figure 2.15: Hydraulic fracture schematic overview from [2] showing multi-stage fracturing (spheres in three different colors) along with a single well .

2.3.1 Microseismic Monitoring

Fracturing is at the core of microseismic monitoring and it is one way to create a reservoir allowing oil and gas trapped in rock pores to flow more easily [70]. This process starts by injecting water or specially developed fluids such as cross-linked gels inside the rock formation. The injection performed under high pressure will lead to cracking the rock and generating the fractures. Furthermore, the generated fracture grows in three dimensions (3D) which propagate creating either a planner or a complex fracture network. Some hydraulic fracturing techniques can be used through multi-stages to increase the drainage area of the well-bore (Figure 2.15 highlights a schematic overview of the microseismic monitoring process showing four different fracture stages along with the injection/monitoring wells). Many seismic events result from this fracturing process and are called microseismic events. They can be used to monitor the fracture growth and image the hy-

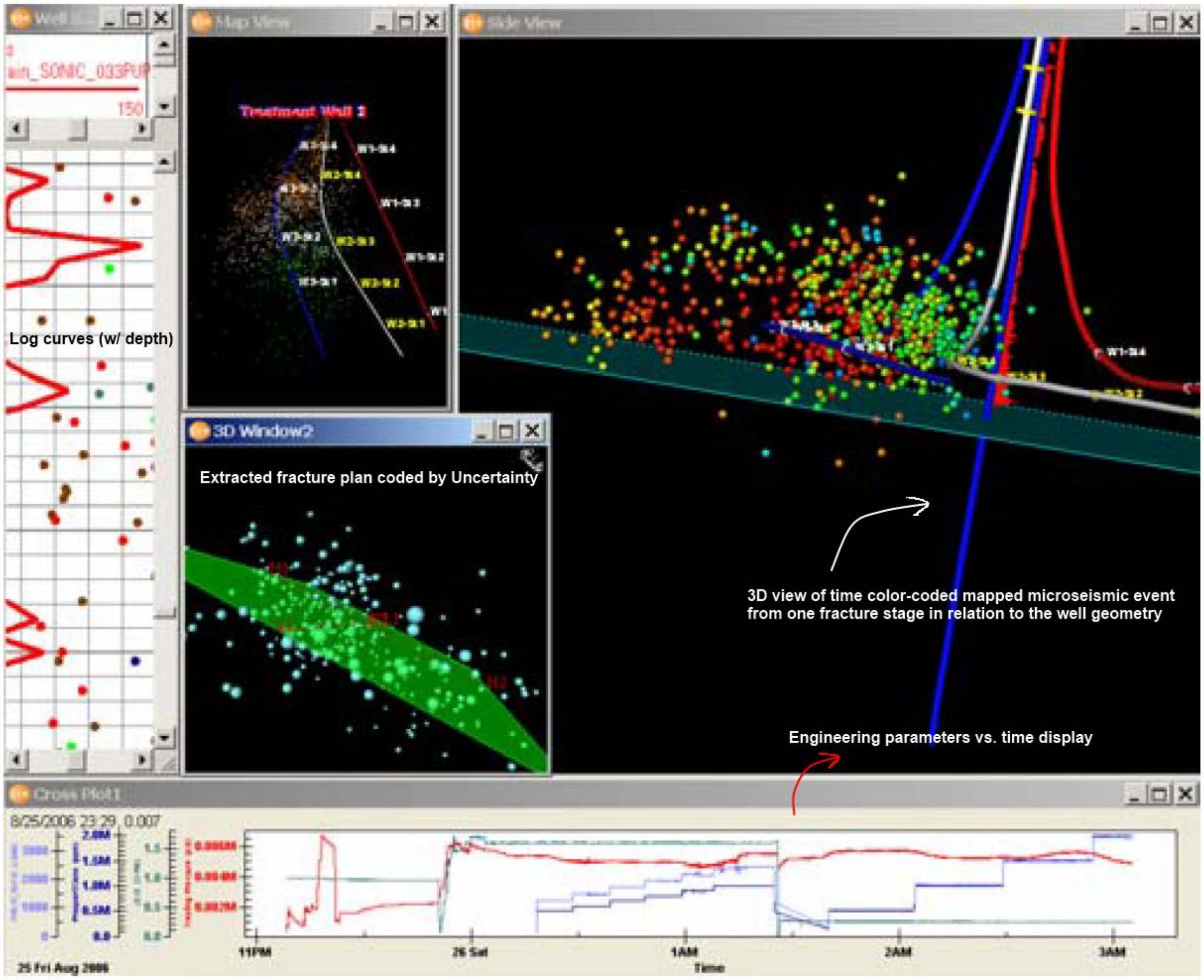

Figure 2.16: Real time microseismic visualization used by Le Calvez [20] illustrating various visual layouts of microseismic data to help the interpretation of the hydraulically induced fracture system .

draulic fracturing dimensions. These events can provide, along with some associated measurements, critical information for optimizing the production. Detection of these events is done by using an offset array of receivers (i.e. geophone sensors) distributed at a specified location near the injection point. The elastic wave generated, which contains a P- and S-wave are used to obtain a spatial location, which consequently map out the geometry of the fracture(s) [94]. However, due to limited acquisition geometry and geologic complexities and noise, uncertainty exists in this model.

Microseismic monitoring has attracted the attention of the oil and gas industry as an important tool in the domains of microseismic engineering and geosciences. Many mathematical methods have been developed to further study this technique [70]. Existing industrial microseismic technology packages such as "Petrel: Microseismic Evaluation" [5] provides many features, tools and plots for working with microseismic events and pumping data. However, most of these tools lack the support of visual interpretation and analysis regarding the high dimensionality of the data and the attributes-correlations.

On the other hand, limited academic research has been done in the area of microseismic visualization. Le Calvez et al. [20] proposed a tool for real-time microseismic monitoring during hydraulic fracturing as per figure 2.16). Tools for seismic interpretation have been created and Marbach et al [57] described multidimensional transfer functions for volume rendering with glyphs as an example (Figure 2.17). Rugis et al. [79] used a 3D reservoir visualization tool for modeling reservoir structures including visualization of microseismic events and their impact on the fracturing process. Amorim et al. [12] applied an interactive sketch-based approach for manipulating the microseismic data along with a developed technique for estimating the stimulated reservoir volume. Figure 2.18 represents visualization and sketch-based interaction to filter the microseismic events intuitively prior to the estimation of the SRV. Moment tensor, a parameter studied in earthquake seismology, can be used in microseismic visualization. For instance, Obermaier et al. [72] use moment tensors to visualize multivariate clustering as shown in figure 2.19. However, the focus of visualizing moment tensors considers only few 3D aspects of the microseismic data and ignore the other aspects so it may not be appropriate for exploring the hydraulic fracture geometry.

2.3.2 Petrography

Petrographic analysis is a technique developed in earth-sciences for observation of microscopic rock samples (thin sections) from cored wellbores. Traditionally, petrography was limited to the identification of rocks and minerals and to the characterization of their properties. Today, however, petrographic techniques are employed to analyze many materials other than minerals; ceramics, glass, to name just a few. Through the process of geological sampling, data is generated and then examined for interpretation. Such interpretation is necessary to describe variations in rock

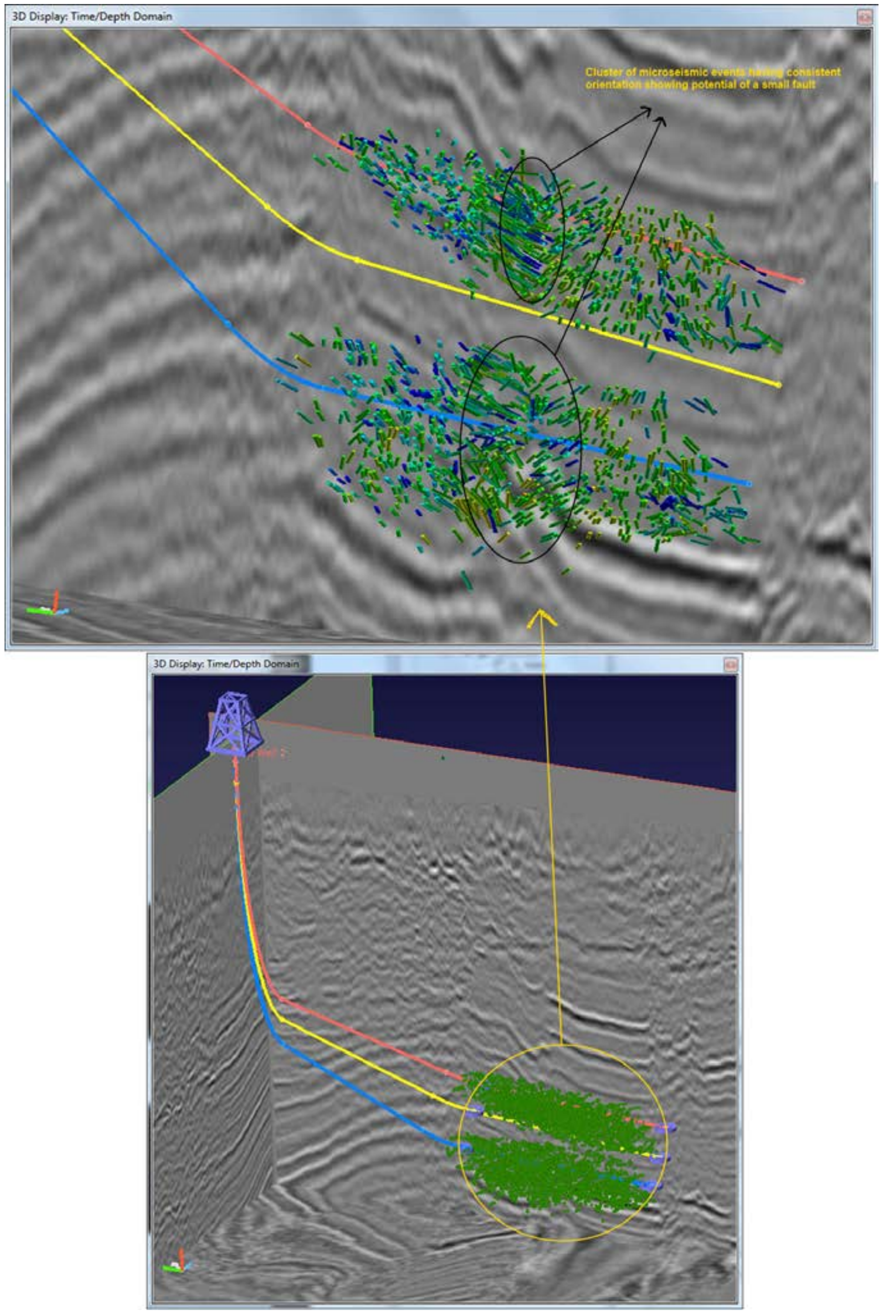

Figure 2.17: Multi-attribute visualization using multivariate volume rendering and glyphs [57].

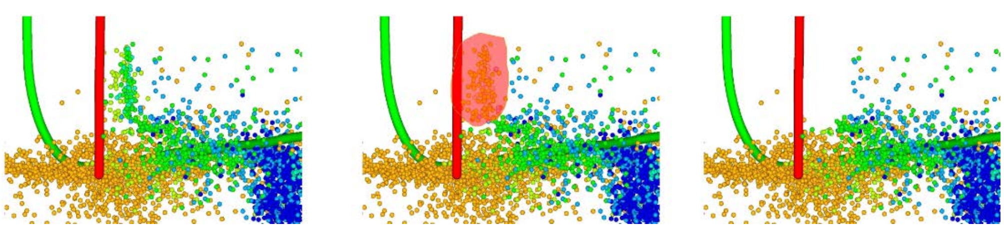

Figure 2.18: Sketching to filter microseismic events while aiming to refine the data subset prior to the calculation of the SRV [12].

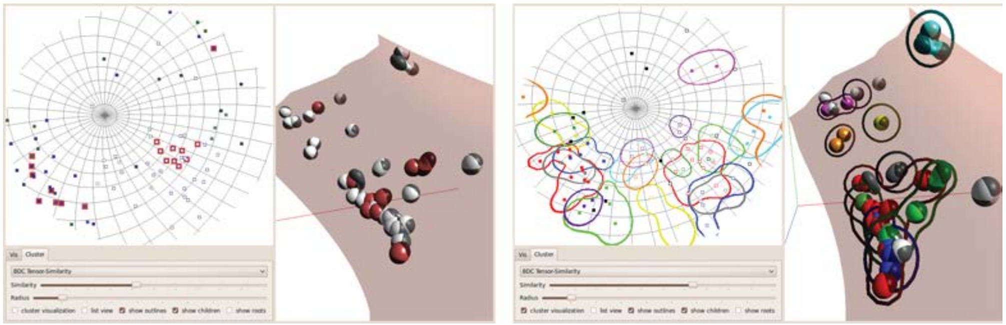

Figure 2.19: Visualization and multivariate clustering of scattered moment tensors, showing a user selection causing some moment tensors to be highlighted along with the identified clusters [72].

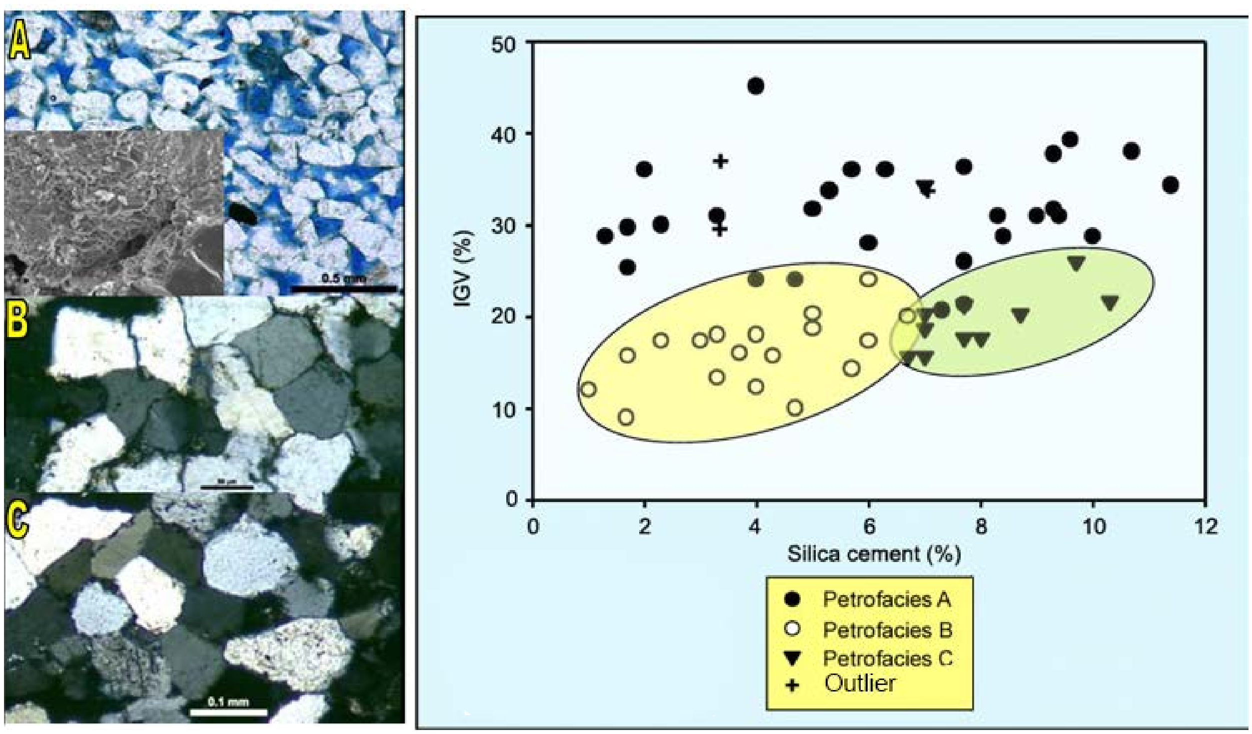

Figure 2.20: Example of reservoir petrofacies (represented in a diagram of some minerals from some rock samples during the characterization) [31].

properties leading to a good characterization of the reservoir. From another perspective, a good characterization of the reservoir would assist in the identification, validation and interpretation of petrofacies [98]. Examples of the petrofacies are illustrated in figure 2.20. Petrofacies represented as a set of petrographic characteristics of the thin-sections allowing the experts to understand the diagenetic processes, aiding in the evaluation of the potential for hydrocarbon storage in the reservoir.

The process of petrographic analysis starts by gathering different rock samples at different depths from one or more wellbores. Each sample is then examined using a specialized (petrographic) microscope in order to learn more about the structure of the rocks, and to identify any minerals within the rock formation. Overview about the major steps involved ni petrographic analysis can is shown in figure 2.21. The results of this examination are produced in tabular databases which classify each sample along with its detailed attributes (also known as compositional data of petrofacies). After that, analysts check these preliminary petrofacies against petrophysical and petrographic quantitative parameters using statistical tools. Finally, reservoir petrofacies can be linked to structural framework parameters for developing coherent models of reservoir quality prediction [31].

Industrial tools for management of reservoir petrographic information such as "Petroledge" (Figure 2.22). Petroledge supports many features to capture, codify, store, process and share the detailed petrographic descriptions of the rocks. Techniques for other types of analysis aims to support macroscopic analysis of the rocks (also known as lithofacies analysis) have been proposed (e.g., [8]). The common approaches of these techniques depend on the recognition of supervised and unsupervised patterns to enable automatic identification of the lithofacies. Furthermore, the majority of these approaches use workflow which integrates methods for dimensionality reduction such as Principle Component Analysis [48], and classification techniques such as support vector machines [91].

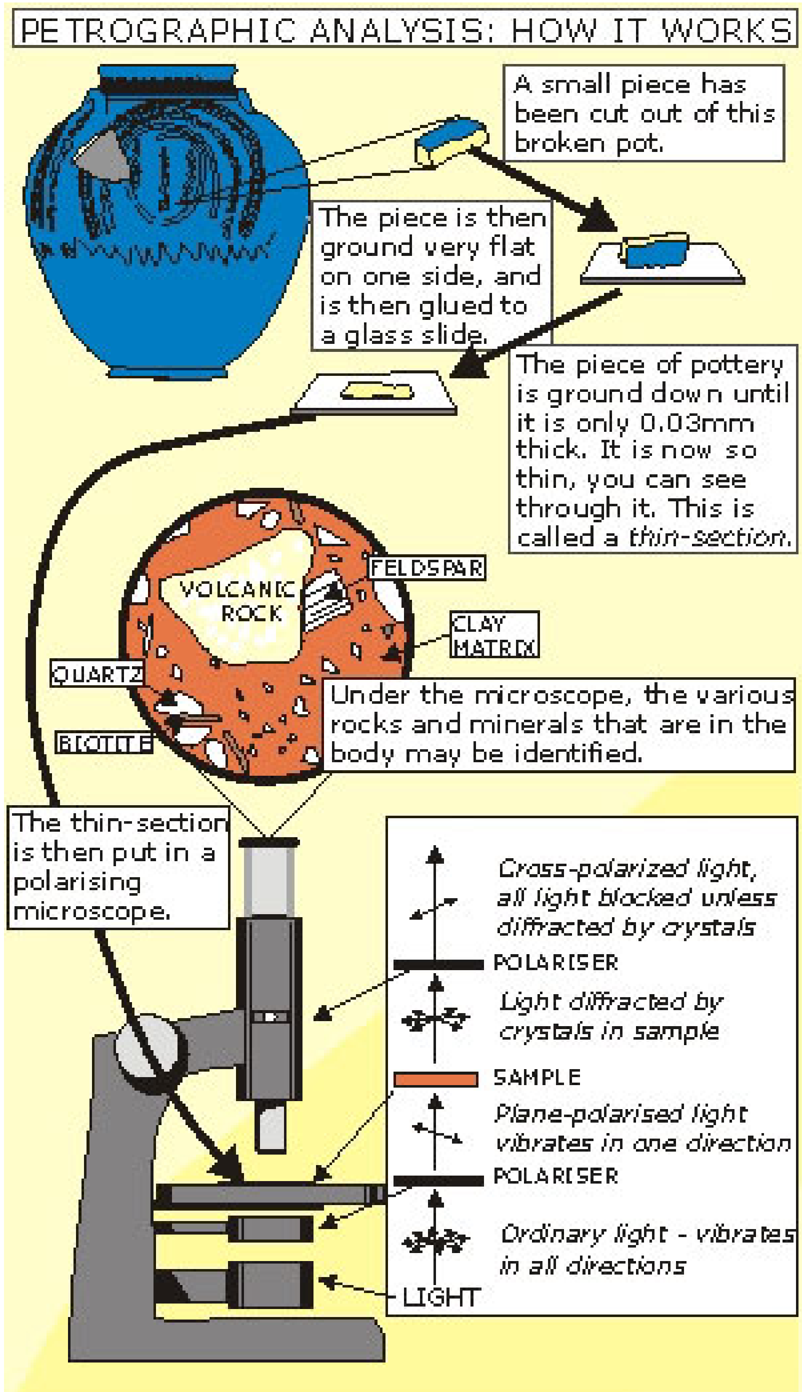

Figure 2.21: Overview of the major steps involved in petrographic analysis [4].

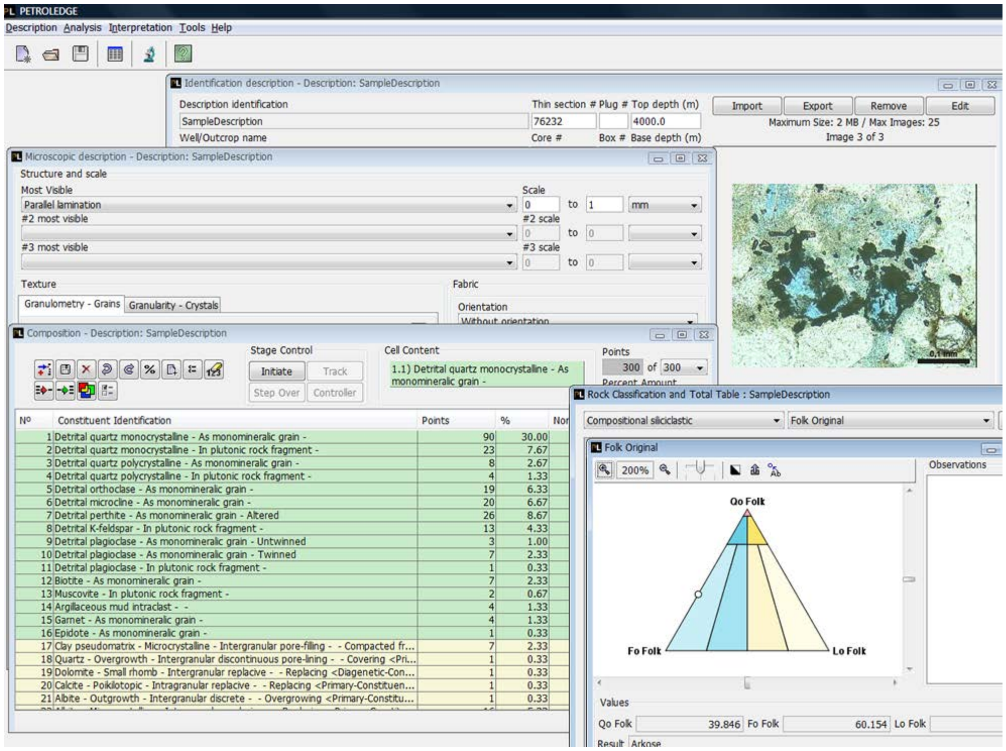

Figure 2.22: Petroledge is a knowledge system for management of reservoir petrographic information [1].

2.4 Summary

In this chapter we discussed the background for this thesis by reviewing key aspects about visual analytics, visualization and interaction techniques. We also presented an overview about two high dimensional oil and gas domain instances representing context for our case studies. In the following chapter, we present our effort involving building a visual-exploratory tool to enable visual analysis of the microseismic data.

Chapter 3

FractVis: Microseismic Characterization and Visualization

Following up on the microseismic background provided in Section 2.3.1, in this chapter we present our efforts in visual exploration of microseismic data. It starts by a characterization of the domain of microseismic monitoring, and follows with a description of a prototype system for interactive visualization and analysis of the multidimensional microseismic data.

3.1 Overview

The primary contribution of this chapter is the characterization of the main challenges faced by the microseismic engineers, outlining the potential benefits of applying scientific and information visualization techniques to this problem domain, and sharing our insight based on the design and evaluation of our current prototype system implementation *FractVis*. We describe the data exploration tasks involved in microseismic monitoring, and the common domain abstractions in order to highlight and share our insight of the domain challenges and needs. From these, we derive our prototype design requirements, encoding choices and interaction techniques. The secondary contribution of our work is the design, development and preliminary evaluation of *FractVis*; an interactive visualization prototype that enable visual exploration and analysis of microseismic events. *FractVis* is being developed and iteratively refined with feedback and consultations from microseismic experts. It combines and extends existing visualization techniques including parallel coordinates (PC), allowing the user to interactively filter and correlate the data during microseismic interpretation and analysis. We describe an illustrative example of the use of *FractVis* in a microseismic monitoring exploration task. This chapter concludes with our reflections and lessons learned during the design of *FractVis*, and directions we propose for future research in applying scientific and information visualization techniques to the domain of microseismic monitoring.

3.2 Methodology

Our approach in the exploration of the domain of microseismic monitoring is exploratory in nature, with sporadic involvement of experts and with exploratory and preliminary results. In fact, while we used unique and real dataset, we were not trying to replace existing tools nor any specific "useful" domain-tasks, since most of these tasks are still being defined.

Our goal is to support microseismic experts in interpreting the high-dimensional microseismic monitoring data. For that, we started collaboration with a group of microseismic specialists. In particular, such collaboration involved meetings with three members of the microseismic industry consortium in the department of Geoscience at the University of Calgary. Two of these members are graduate students, while the third one is a professional with previous industrial experience and now is the group leader of the consortium. The collaboration structure was ad hoc as we only met whenever needed, around once per month (on average) for the duration of one year, and each collaboration session lasted around one hour. During the early meetings, the domain experts provided feedback regarding their work processes, challenges and needs, and also regarding the domain jargon. Later meetings focused more on giving feedback about our visualization prototype. The raw data collected from these meetings were in the form of notes comprising: (1) annotated sketches of different aspects regarding the microseismic monitoring process and (2) key statements of the experts clarifying important domain concepts. Some of the collaboration sessions were recorded to gather feedback from the experts, and to allow us to easily access their explanations regarding the domain as well as regarding our tool whenever needed. In particular, one of the meetings was audio recorded where the expert detailed some feedback about our visualization prototype, and clarified the domain's expectations. Five other sessions were video recorded, with two of them involved observing an expert while doing her work, and the remaining sessions contained the experts' feedback regarding our visualization-tool. In general, our approach of collecting the raw data was inspired by observational techniques such as contextual inquiry [44] and rapid ethnography [66].

We also received a real microseismic dataset from our domain collaborators. Different aspects regarding this dataset are outlined and explained in Section 3.3.1, including a description of the high-level tasks that the experts typically perform in their analysis. Our goal was to focus on building an exploratory visualization of the given dataset to assist the experts in interpreting the high-dimensional data.

Toward achieving our goal, we developed a visual exploratory prototype with different visualizations and interactions. At the early phase of the development, different ideas were implemented while exploring how to visualize the data aspects (Section 3.5.4). Our design choices as well as the developed prototype have been refined and improved because we were gaining a better understanding of the experts' requirements after each collaboration session. For example, during one of the video-recorded sessions, one expert gave a suggestion about the visualization of shadow boxes, and the same expert in a later session provided her feedback about our implementation of that feature. While considering our approach as being iterative, yet we did not have a clear comprehensive understanding of the domain and its complexities. In fact, we did not have guidance through a clearly-defined real-task while developing our visualization. Therefore, our approach should be considered more of ad-hoc exploration.

The final outcome of our exploration of the microseismic monitoring domain was a simple characterization of it and a development of an exploratory visualization prototype with the potential for supporting the visual analysis of the high-dimensional microseismic data. During and after the development, we received feedbacks about the visualization and the interactions. However, due to time and limitations, we did not perform a full formal evaluation, we had only a general qualitative feedback and suggestions for improvements.

3.3 Microseismic Characterization

Microseismic monitoring offers unique scientific and information visualization challenges and potential. In this section we attempt to briefly characterize the microseismic monitoring domain in order to motivate our own design, and in hope that this characterization would allow future visualization efforts to address the various domain challenges. We describe the typical structure of microseismic monitoring datasets, and highlight its important attributes. We also present the data abstractions experts are using when approaching the datasets and the high-level tasks they are pursuing, along with the processes and the challenges they are facing. The raw data we present was gathered during continuous meetings and consultation sessions with domain experts and collaborators, contextual inquiry [44] sessions, and semi-structured interviews to learn more about microseismic monitoring practices and needs. During the sessions, we employed a questionnaire we designed to find out and confirm our understanding about the requirements and needs of the domain experts (a summarized version of the questionnaire, and a set of sample answers are available in Appendix B).

3.3.1 Data Description

Microseismic dataset is composed of many layers. In our work, we focus on three layers which are related though time and space. The first layer is the microseismic “Events Catalog” which describes each event along with its attributes. The second layer is the “Monitoring and Treatment” wells information. A third component comprises the engineering data and pumping curves. All of these data layers usually exist within a single microseismic dataset. In this part, we describe in details the structure of the microseismic data events (the first layer) along with a simplified explanation of the most important attributes of each event. We also provide an overview of the engineering data (the third layer) and its importance. However, we do not describe specific information about the wells, since it represent their fixed locations. The microseismic data employed in our design of *FractVis* was a multidimensional, time-varying point cloud dataset. This dataset is the result of a first-pass preprocessing of the raw microseismic signals collected in the field during microseismic event monitoring. This point cloud dataset has more than five-thousand events, with 35+ attributes per event. The distribution of these events relates to different microseismic fracture stages. Following is an overview of some of the important attributes of microseismic events:

1. **Time**: hour, minute and second; this set of attributes refer to the time the microseismic event occurred.
2. **MS-LOC-X, MS-LOC-Y, MS-LOC-Z**: The spatial location of a microseismic event (in meters).
3. **MS-DISTANCE**: Distance from the event position to the monitoring sensor; It can be calculated by determining the time delay between the Primary-and Secondary-wave arrivals.
4. **MS-LOC-SNR**: Signal-to-Noise Ratio, measures an estimate of how much the signal has been corrupted by noise. It can express the uncertainty of the data being measured.
5. **NOISE-LEVEL**: Level of noise in the microseismic event.
6. **PSH-AMPL-RATIO**: Ratio between P-wave and S-wave amplitudes.
7. **SP-RADIUS**: determines the size of the area ruptured by the event.
8. **SP-MOMENT**: the seismic moment; represents the scalar measure of an earthquake rupture-size related to the action of force across the area of the fault.
9. **SP-MAGNITUDE**: a log scale, similar to Richter scale for earthquakes.
10. **SP-ENERGY**: energy is calculated by considering the history of a particle as it responds to a transient seismic wave field. As a wave passes, the particle, which has a potential energy, will have a velocity and thus a kinetic energy. The sum of the potential and kinetic energies, integrated over time, will yield the particle energy.

The insight shared by domain experts suggests that some of these attributes are more important than others. They also expressed that some of these attributes are independent while others are dependent of each other. For example, the attributes *distance* and *magnitude* are independent and are usually used as standard test for initially checking the validity of the data. The attribute *noise level*, on the other hand is dependent on the *signal-to-noise ratio* attribute.

Figure 3.1: Example of the microseismic engineering (pumping) curves [6]

The engineering data layer represents the different characteristics of the fracture growth and the events population with time (Example of such engineering curves is shown in figure 3.1). For example, pumping curves provide correlation between time and pressure in the injection process. By examining these real-time plots, experts can confirm that microseismic events start to be generated when the pressure reaches its peak with the fluid injection causing the rock to break or fracture. In addition, they can also see that the events continue to populate while chemical and fluids are being injected to hold the fracture open. Visualizing the engineering data layer curves and linking them with 3D visualization of the events is important for better understanding of the fracture geometry. The current version of *FractVis* does not address this second layer of the dataset and we are planning to integrate it in our future prototypes.

3.3.2 Challenges

Attempting to analyze the microseismic monitoring dataset involves several challenges. First, although some of the data attributes have dependency, the dimensionality of the independent data attributes is still quite high. Potential techniques for reducing this high dimensionality would certainly aid in the analysis of the data. Second, the data inherently contains uncertainty due to the inaccurate measurements and the noise associated with them. Noise in the data comes from many sources and cannot be completely removed. Although many techniques for reducing the noise have been attempted recently [90], the processed data still contains a noise quantification information associated with each event. This noise quantification is being reported through the attributes *noise level* and *signal-to-noise ratio*. Finally, the microseismic data is highly abstract. The data could have different interpretations and it can often be difficult to validate which of them is the most accurate one. Experts explained that some of the attributes may have different meanings in different contexts, and that applying domain insight is still a crucial part of the process. Adding to this, there is a lack of computational tools that support intuitive interpretation of microseismic geophysical data, and most of the current tools are either too general/expensive or in-house propriety solutions. Overall, the domain experts we consulted thought that visual analytic tools would be very effective in helping them to explore the dataset interactively and effectively. For instance, domain experts reported they sometimes do not fully understand the relationship between many of these attributes and were hoping to be able to intuitively spatially correlate various data attributes in order to learn more about the potential effect of each of them.

3.3.3 Data Abstraction and Task Analysis

Microseismic experts perform different tasks while exploring and analyzing the data. In this section we describe the most important ones. First, domain experts reported that estimating the Stimulated Reservoir Volume (SRV) is one of the most common and important tasks in microseismic engineering. The goal of this task is to generate a bounding volume which defines subsets of the data events as initial estimations of the production volume. The locations of the events are important in this calculation. However, these locations are estimated due to measurements that inherit uncertainty. Experts consider the inclusion of uncertainty in SRV calculation as an important future challenge [90]. Various methods are applied to the events prior to calculating the SRV in an attempt to filter and analyze the data attributes. In this task, the expert needs to focus and select from the subset of the events the ones that are considered to be the important ones for estimating both the SRV and the fracture growth. The ability to filter the data and make decisions regarding the events is greatly affected by the insight and understanding of the dataset, and the expert's ability to extract relations among its attributes. Additionally, experts are analyzing dataset outliers manually. They believe that this manual component can benefit greatly from applying (semi-automatic) interactive visualization correlation techniques. Secondly, since the microseismic data is a time-varying point-cloud, there is room for supporting time-based visualization and analysis. Microseismic experts consider the *time* attribute to be one of the most important independent variables. They expressed that it is common to analyze the correlation between the *time* and many other attributes. Thirdly, analyzing fracture growth over time (i.e. measurements of fracture azimuth, width, etc.) could be spatially visualized to obtain an understanding of fracture geometry and the fractures' interactions, understanding that can be crucial when analyzing the dataset. Finally, domain experts expressed that the ability to see the data from different perspectives at the same time is important. For instance, the synchronized: visualization of the 3D events, visualization of the attributes, and the visualization of the engineering curves would be useful if represented intuitively.

3.4 Design Rationale

With the implications resulting from the problem characterization, we decided to design and implement a prototype to allow interactive visualization and analysis of microseismic data. We decided to focus on supporting the simple tasks of "data filtering" and "attributes correlation". We adopted an iterative design approach, where we built our first prototype, and modified our system iteratively based on the requirements and the feedback we received directly from domain experts and collaborators. In this section, we describe and justify our design decisions; which concrete tasks to support; the visualization requirements and which technique to use and/or extend; and the best encoding for presenting the information. We also discuss our choices regarding the data representations and the possible interaction mechanisms.

We analyzed the high-level tasks of "data filtering" and "attributes correlation" then we identified the following concrete tasks: *Find Anomalies*, *Associate*, *Correlate*, *Identify*, *Filter*, and *Categorize*; by following the taxonomy of [11]. As a result, we designed our prototype to support these tasks.

In the next two subsections, we outline and discuss our visual encoding choices and the possible representations that could be used for visualizing microseismic data.

3.4.1 Visual Encoding

Some visualization can misinform people because of weakly designed data representations. Therefore, we carefully combined the use of mapping of visual variables with our iterative discussions with domain experts in order to best insure that the data is clearly represented as visual elements, in a way that would clearly emphasize the importance of certain data attributes [84]. Furthermore, we created various (design) sketches to jot down different ideas and concepts while exploring the possible representations.

We chose to represent every microseismic event as a sphere whose center is the 3D spatial location of the event, and to be able to map the sphere's radius to be proportional to any of the event's attributes. Regarding the choice of colors, we decided to color the spheres as well as the PC's content according to some arbitrary color map. In other words, the color of each sphere is defined by relating one of the event's attributes to the active color map. Among the color maps that we supported, is the rainbow (jet) color map, which may not be recommended for usage in visualization systems [17], but domain experts are familiar with this color map. So then, we think it is natural for the domain expert to see different (distinct) shades of colored spheres according to the different categorization/distribution in the data. Our domain collaborators acknowledged our choices of mapping the radius/color of each event sphere relative to some attributes. They considered this mapping to be natural to them, powerful for showing much information at once, and comparable to many existing commercial tools.

3.4.2 Why Parallel Coordinates?

It is worth mentioning that we did not explore other multi-dimensional visualization techniques, and we are not sure if other techniques would be a better choice or not. In fact, we wanted to focus on the PC technique, and below are a set of reasons supporting our decision.

Different reasons supported our choice of the PC technique to visualize the multidimensional microseismic data. First, the technique of PC supports exploration of data trends and correlation of the attributes without affecting the scale and the dimensionality of the data, which is not the case for the other projective and non-projective techniques. Therefore, we consider it a good fit for our data. Second, PC is a widely used technique and supports extensibility. Indeed, we extended the basic implementation of the PC by integrating dynamic magic lenses and embedding them with it. Furthermore, experts can dynamically recolor the content of the PC according to some attribute to examine attributes correlation without the need to perform axis reordering. Third, the study performed by Siirtola and Räihä [86] revealed that people who performed their tasks with PC found it more effective than those who used some other methods. In general, domain experts are mostly familiar with plots from scientific tools such as *Matlab* [3]. We believe that it would make more sense if we provided them with a simple 2D plot – enhanced PC in our case – similarly to what they expect. Finally, we proposed that if we extended our visualization and provided interactively embedded visuals (e.g. scatter plot) within our PC, then it would be easier for the experts to familiarize themselves with it and learn interacting with it quickly, interaction which would empower them with rich visuals without the need to open additional visualization windows.

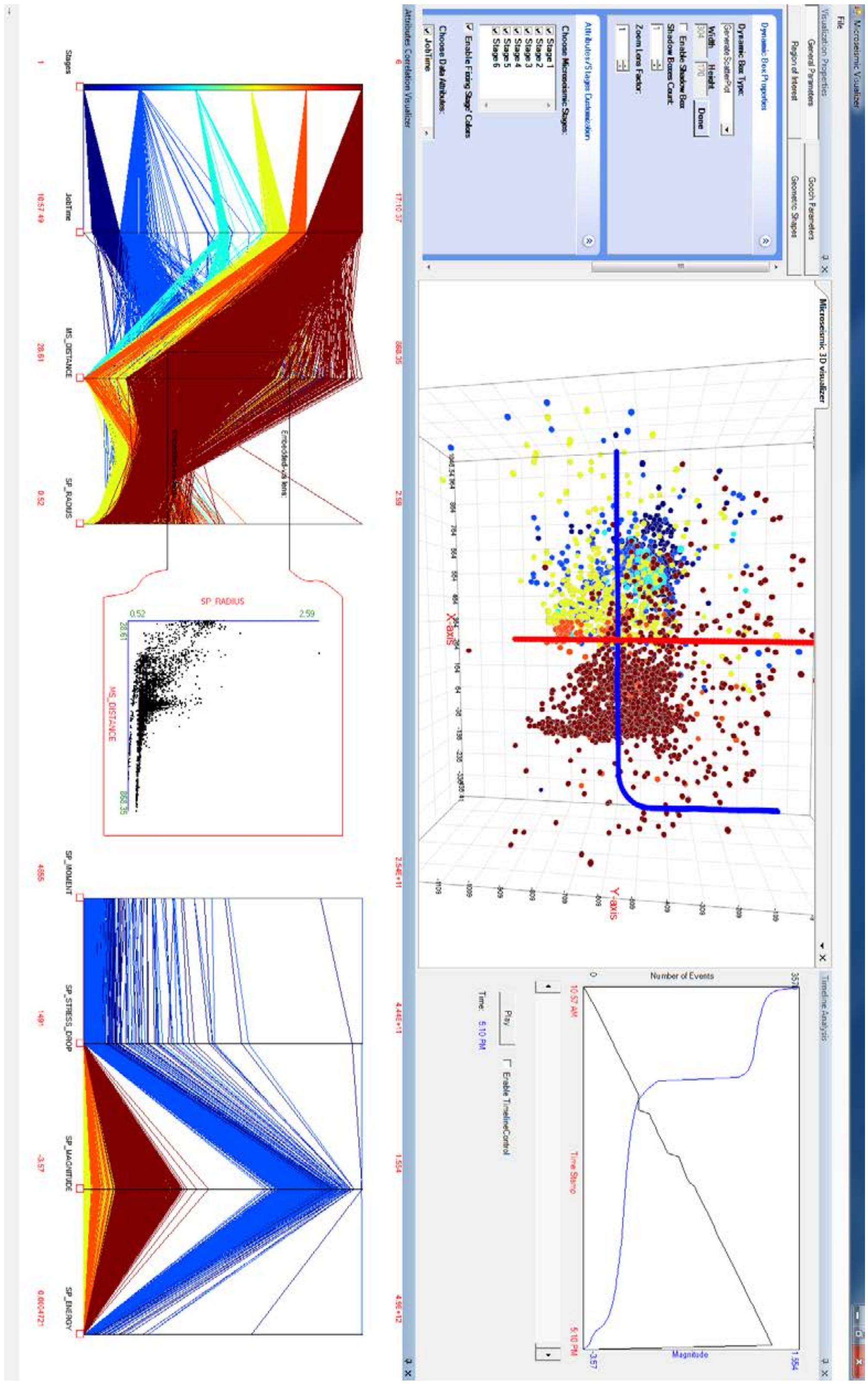

Figure 3.2: System overview showing the synchronization of the PC view (bottom) with the other data visualization components: (top center) 3D microseismic events' point cloud, (top right) the time-based visualization and (top left) the GUI view for controlling the visualization parameters.

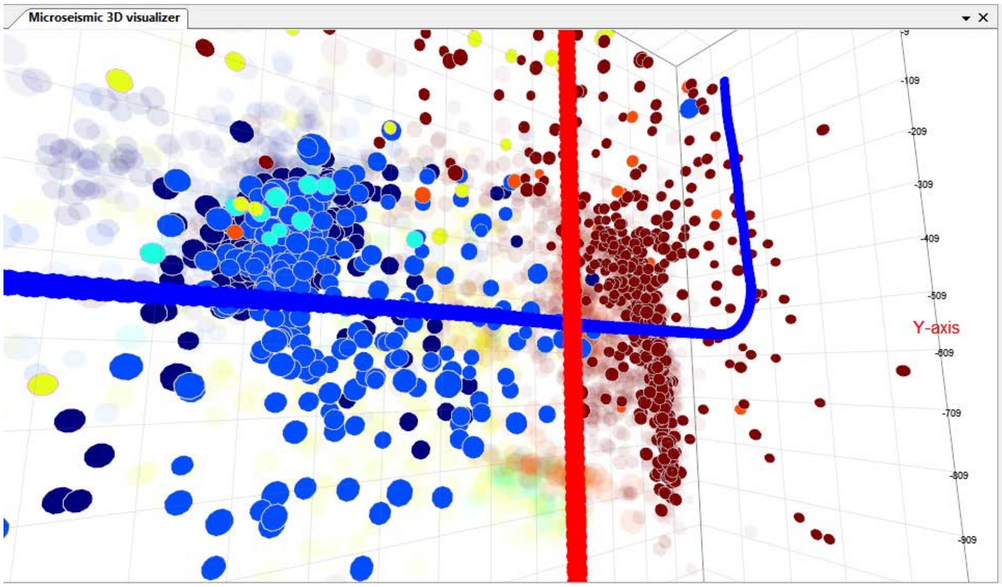

Figure 3.3: 3D visualization view showing the visible events being rendered (solid with their outline) using rainbow (jet) color map. The filtered out events are rendered transparent (without outline) to keep the context.

3.5 FractVis

Our implementation follows the multiple-coordinated-views approach. We considered this approach as we observed, from discussions with domain experts and collaborators, that it is important to have different representations of the data at the same time in order to provide visual cues and insight to facilitate simultaneous data analysis. Our system, *FractVis*, supports three coordinated visualization views. In Figure 3.2; the main 3D view enables exploration and visual analysis of the microseismic events in the reservoir space with well integration (Section 3.5.1), the second view supports flexible interaction and correlation though an improved PC visualization (Section 3.5.2), and the third view aims at supporting time-based analysis of the data (Section 3.5.3). Each view presents (some aspects of) the data in a different way, allowing people to link and relate the meanings gained from one view with the others views. Finally, we have a specialized view for controlling the visualizations parameters.

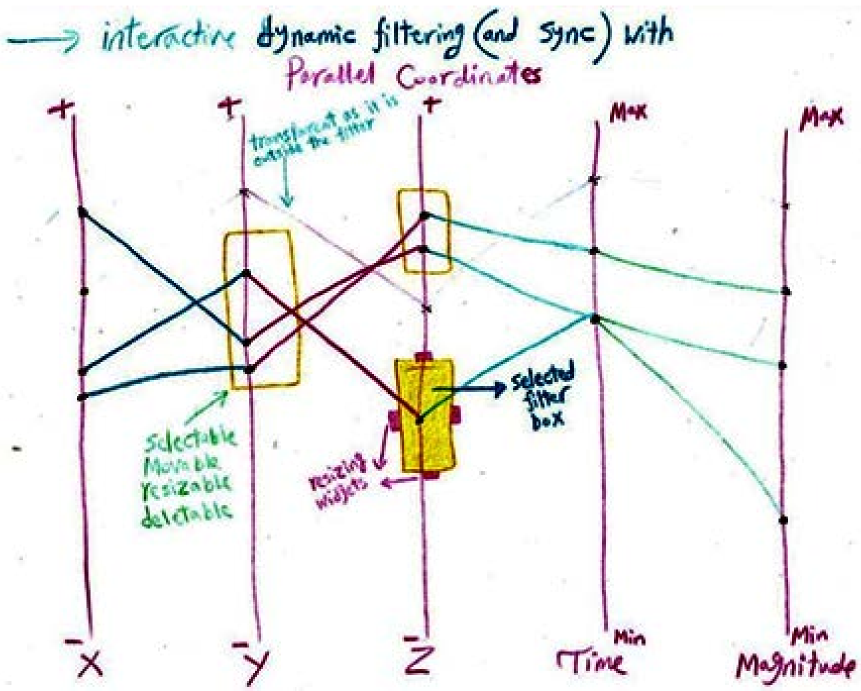

Figure 3.4: A sketch showing our initial visualization of parallel coordinates with the integration of dynamic filter boxes.

3.5.1 3D Spatial Visualization

The main goal of this visualization approach is to represent the data in its spatial distribution, providing basic insights about the microseismic geometry. This is mainly useful for 3D exploration and correlation in order to better understand the geometrical behavior of the microseismic events. For example, the 3D visualization would support (1) the analysis of the distribution of the events around the (treatment) well, and (2) the understanding of the hydraulic fracture shape (growth). This 3D visualization window displays every event as a sphere, where the sphere's color-and-radius encodes an attribute value (Figure 3.3). In the example shown in figure. 3.3, the surface's color of each sphere is relative to the event's stage number. The visual variables (color, size and transparency) that have been used in this view are updated according to any interactions over the PC. Furthermore, we supported different manipulations that can be applied over the 3D events including the ability to manually remove any event by simple mouse selection.

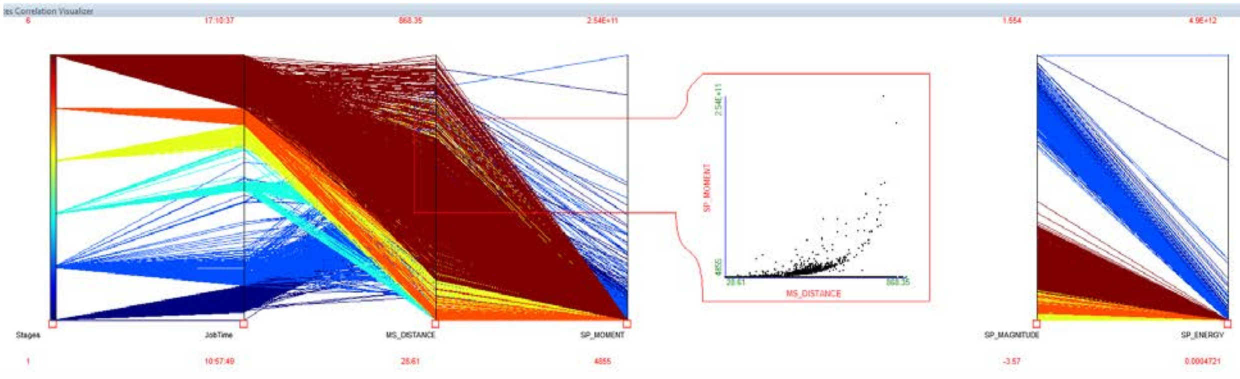

Figure 3.5: Overview of the parallel coordinates visualization inside FractVis highlighting color-based correlation and the embedding of scatterplot within the visualization.

3.5.2 Parallel Coordinates Visualization

The visualization technique of PC [46] can be used to visually explore and discover hidden relations of a multidimensional data. It is considered a robust way to visualize high-dimensional geometry and support analyzing multivariate data. The standard PC consists of n-parallel lines typically vertical and equally spaced, where 'n' is the number of dimensions (attributes) of the data. Each data sample is represented as a polyline intersecting each attribute at the corresponding relative value.

We created different sketches to investigate and explore the best way to adapt the PC technique for visualizing the microseismic data. At first, we explored the idea of integrating dynamic filter boxes combined with different coloring possibilities into the PC visualization (an example of such sketches is shown in figure 3.4). Figure 3.5 shows our implementation of the parallel coordinates within FractVis along with a set of interactions.

We extended the PC by introducing and integrating two novel extensions with the hope of improving visualization and interaction of the microseismic seismic data. The first extension describes the integration of magic lenses over the PC with the goal of providing intuitive interactions including data filtering and scaling. In the second extension we present our idea of visual correlation through the use of visual legends.

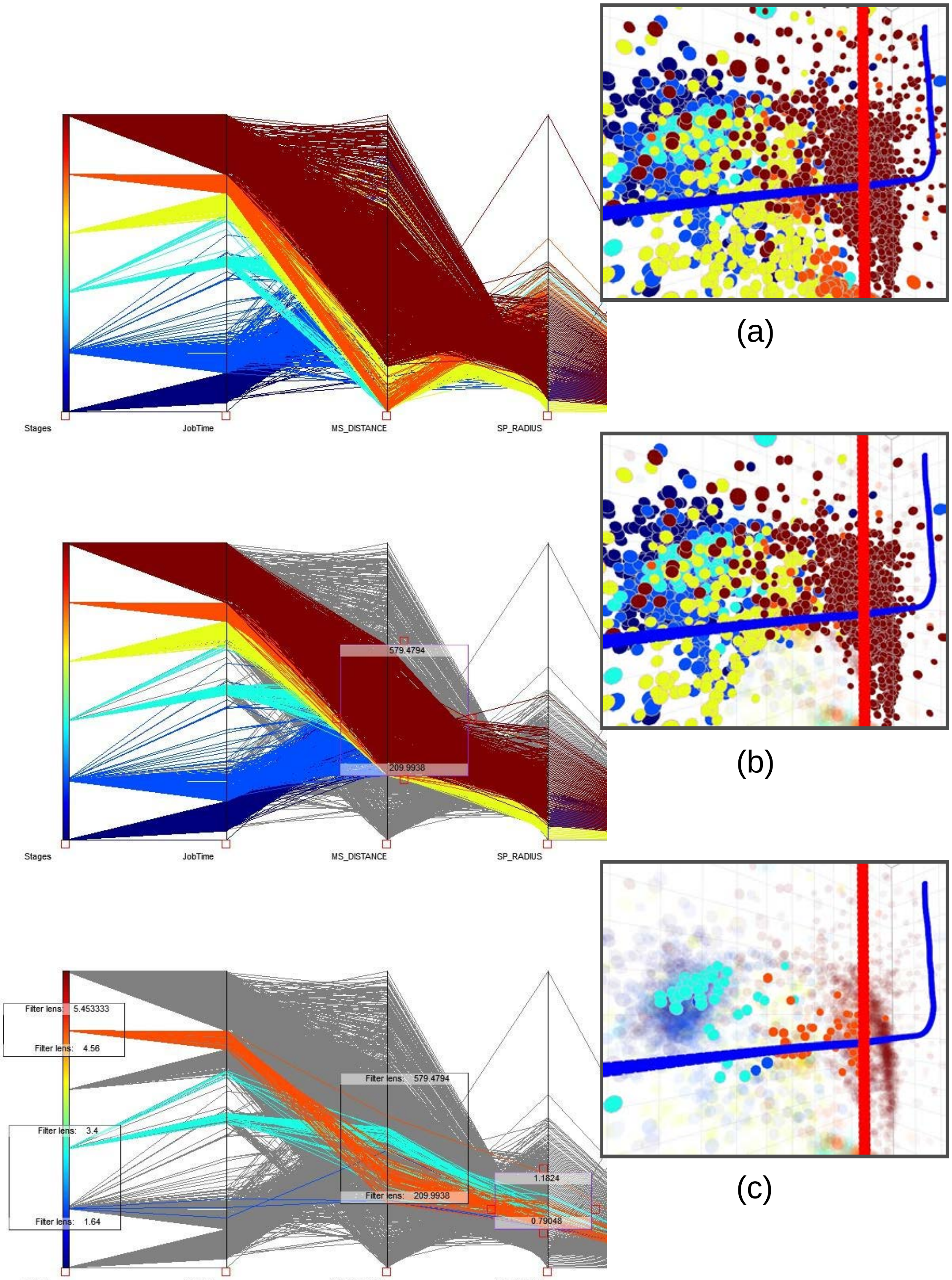

Figure 3.6: The effect of one dynamic box (lens) which filters the PC data as well as the visualization in the 3D view. All the filtered out PC's samples are being shown in gray lines (and transparent spheres) to keep the context. Here, (a) shows the visualization without any filtering, (b) shows the effect one filter box, and (c) shows the effect with many filter boxes.

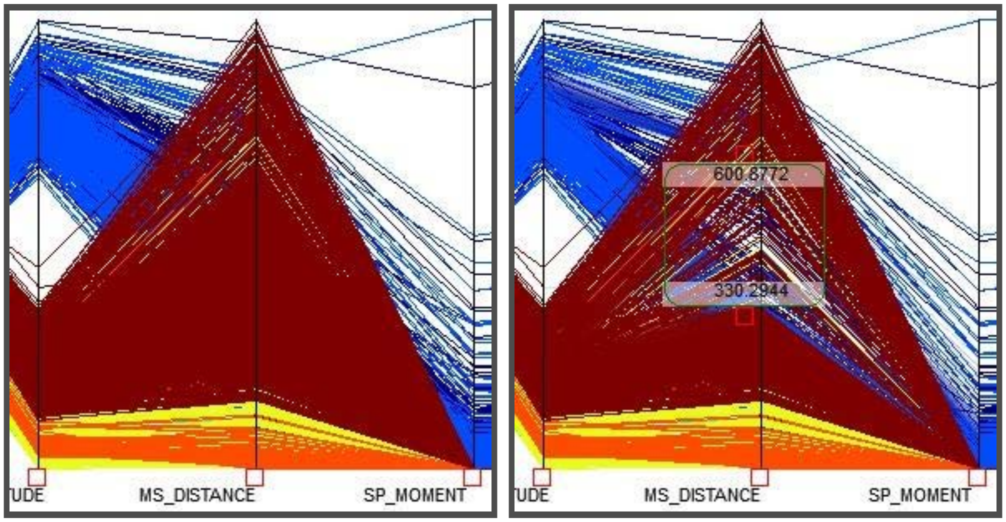

Figure 3.7: The effect of one dynamic box (lens) which scales the PC range to reducing cluttering.

Magic lenses over parallel coordinates

We extended the implementation of the PC through the concept of dynamic boxes (similar to magic lenses [16]) blended over the PC's plot (Figure 3.4). We support integration of different visuals (by embedding them) according to the type of the dynamic-box being used. For instance, the user can show scatter plot within the PC visualization if a dynamic box reflecting such integration has been used. While Holten and van Wijk presented that idea before [43], in our implementation we support such integration interactively. Once a dynamic box is created, all the visible attributes' axes that intersect this dynamic box will be considered for achieving the corresponding effect. We support three different dynamic boxes each one of them is represented using different color and shape to utilize the cognitive power of the users and to facilitate interactions. Following we describe each of them.

The first dynamic box causes data filtering (Figure 3.6). Such a filter box will constrain the events to only those who fall within its limits (range), similar to data brushing [62]. For example, the filter box shown (in Figure 3.6 (b)) which is being placed over the *distance* attribute, shows only the events that have certain distance values. Furthermore, our visualization shows the filtered out events as transparent 3D spheres and gray polylines within the PC to make it easier to identify them. The user can create many filter boxes to achieve complex filtering. This idea is similar to iterative brushing [37] which enables the creation of composite filters to refine the data subset.

The second type of the dynamic boxes is a box that enables integration of custom visuals within the PC visualization. Figure 3.2 (bottom) shows a dynamic box that caused a scatter plot's visual to be generated (and embedded) within the PC.

The third type of the dynamic boxes aims to reduce data cluttering similar to the work of Ellis et al. [35]. The box causes zooming or scaling each attribute's range to reveal the density of the data elements (Figure 3.7).

Shadow boxes (Figure 3.8) are other novel visual elements that can be attached with filter boxes to enable: (1) range/cluster navigation; by gradually fading all the events before and after the range of the current active filter box, and (2) partial contextualization. The number of shadow boxes as well as their properties can be controlled through a specialized Graphical User Interface (GUI) panel. For example, in Figure 3.8 (b and c), the effect of shadow boxes is being shown. We can see that although we are strictly filtering the data events (using our filter box), the (synchronized) 3D view shows other (transparent) events representing the partial context. We can also notice that the transparency of the shadow boxes (and the events) increases gradually creating a partial context and facilitating clusters navigation. Furthermore, the contour of all the visible events is rendered to enable quick identification of the unfiltered events among the other transparent (filtered) ones. In essence, The idea of shadow boxes aims to intuitively show the immediate neighboring context of the current active filter box, representing an extension to the technique of focus+context [23].

Visual legends

We introduce a new way to perform dynamic correlation of the high-dimensional data attributes through the concept of "visual legends". In other words, the visual legends aim to provide experts with the ability to perform easier correlation (1) between the data attributes, as well as (2) between the data representations inside the PC visualization and the other linked visualizations (e.g., the 3D view). The basic idea is about placing visual maps (i.e. color map) over any attribute's axis

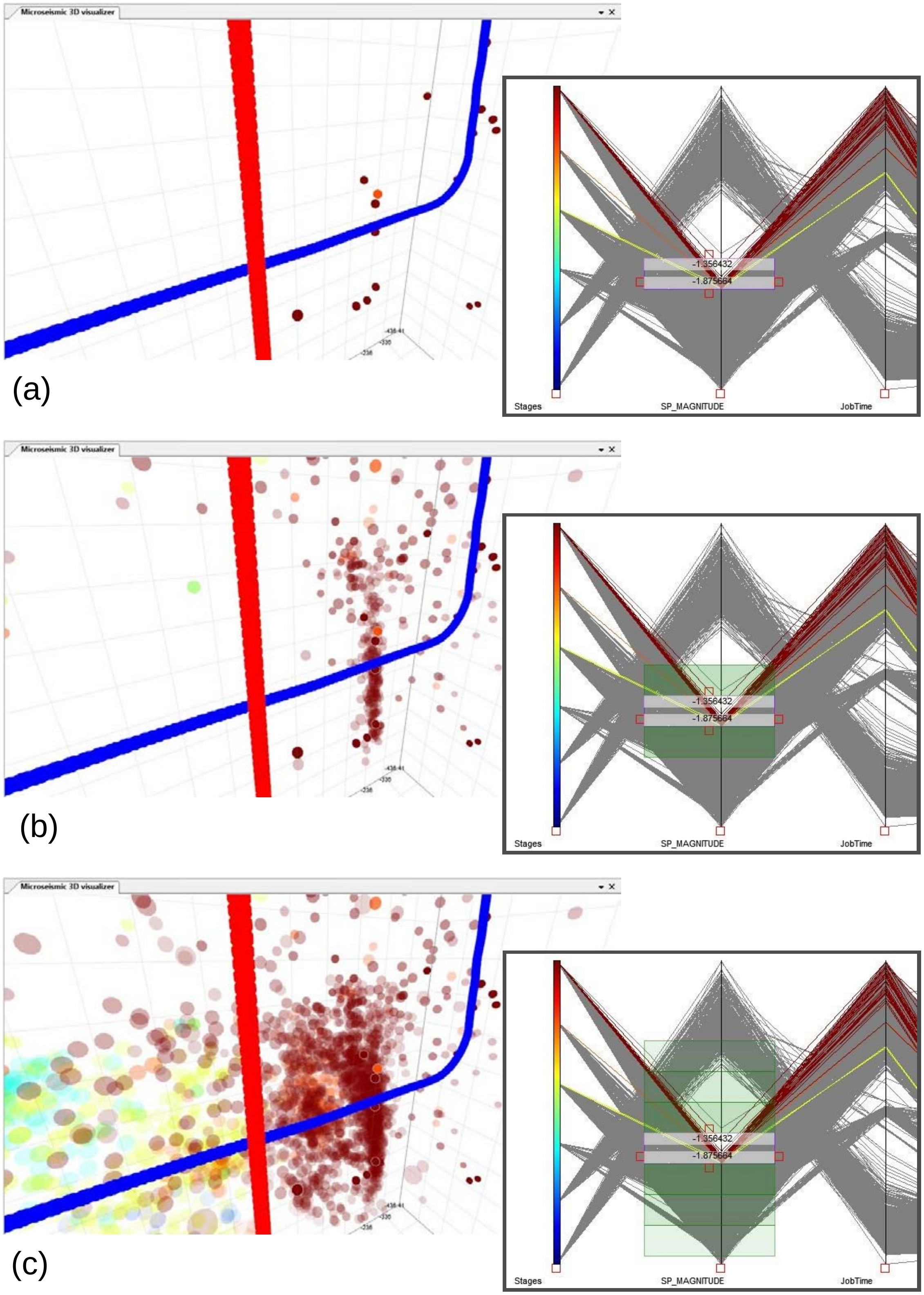

Figure 3.8: The effect of shadow boxes over the PC: (a) show the normal filtering with a single filter box, (b) show the effect of two shadow boxes and how they result in a small partial context, and (c) show the effect of activating six shadow boxes to increase the partial context.

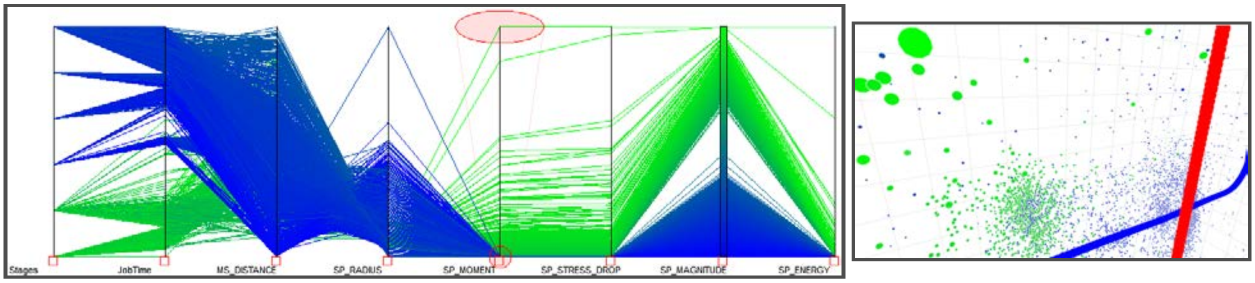

Figure 3.9: The effect of applying two visual legends (maps) over our PC: The “green-to-blue” color map (over the *magnitude* axis) which recolor the PC’s lines & the 3D spheres according to the *magnitude* attribute, as well as the size map (over the *moment* axis) which resize all the 3D spheres according to the corresponding *moment* values.

to update the data representation according to that attribute. This idea is similar to “gradient color brush” introduced by Matkovic et al. [63] but we extended this idea by allowing it to represent different visual variables such as color and size. In our implementation, we support two visual legends (maps) in order to perform color-correlation and/or size-correlation (Figure. 3.9). First, a color map can be placed over any attribute to (associate and) enforce (re)coloring of all the PC’s polylines, as well as the 3D spheres, according to the distribution of the values of the selected attribute. We supported two color maps (green-to-blue and rainbow) that can be used as color legends. The example shown (Figure 3.2) depicts a rainbow color-map blended over the event-stage-number’s attribute causing the visual elements of each event to be colored according to the corresponding stage number. Second, a size-map can be placed over any attribute to (re)size all the sphere-events accordingly. This could help in identifying the spatial geometric location of events relative to the well. Furthermore, it can be also useful for analyzing the 3D location of any possible events’ outliers and confirm if they are outliers or not. For example, placing this legend over the *moment* column enables the association between the event sphere’s size with the event’s moment value; the higher the moment of an event, the bigger its sphere radius (Figure. 3.9).

Our implementation also supports the feature of axes reordering to analyze the correlation between any non-sequenced attributes, but we believe that our legend-based correlation could enable such correlations without the need to reorder the attributes, thus leading to faster data analysis.

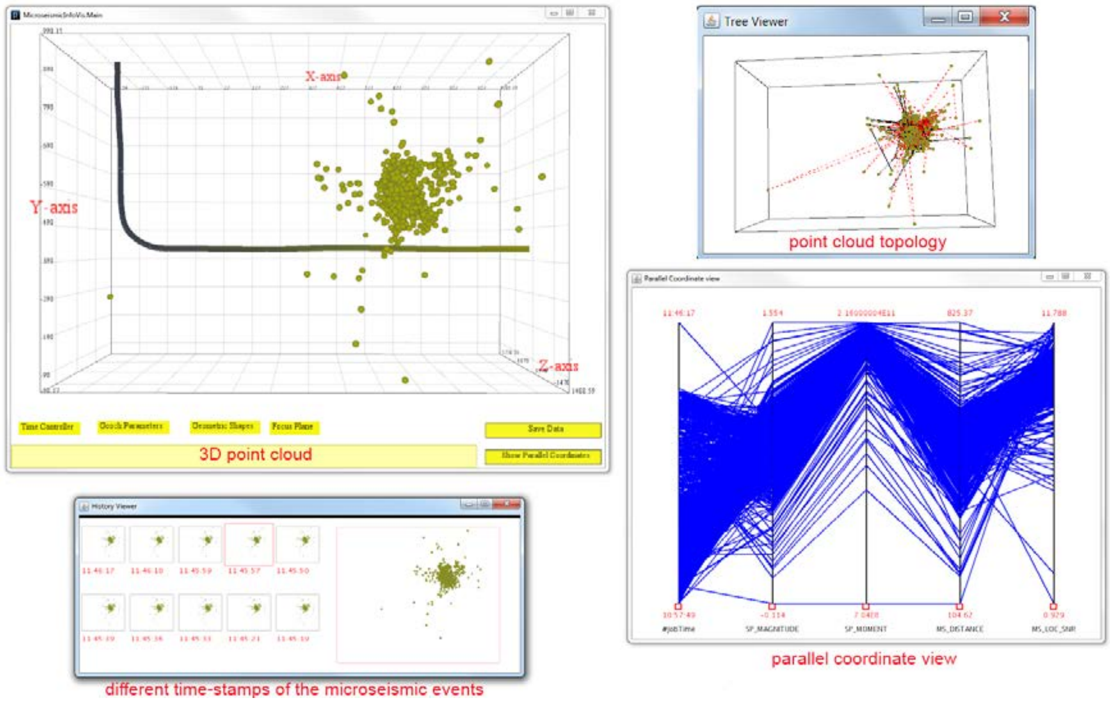

Figure 3.10: Synchronization of the parallel coordinates view (1) with other data visualization components: (2) 3D point cloud, (3) different time-stamps of the microseismic events and (4) the point cloud topology.

3.5.3 Time-based Visualization

Microseismic data is a time-variant dataset and the *time* attribute of the data is important for many analytical operations. We implemented a basic time-based visualization view (Figure 3.2 top right). Through this visualization, the expert can perform basic time-based analysis including examining the total accumulation of events with time. Moreover, the expert can use a time-slider in order to filter and navigate the events proportional to certain time-step. For instance, to show only those events that have been generated after some time range value. The user can use the "play" button to animate and simulate the population of events with time. Finally, we believe that extending *FractVis* with additional time-based visualization features would be of particular importance to the visual analysis of the engineering (pumping) data. Indeed, we are considering this extension as part of our future work.

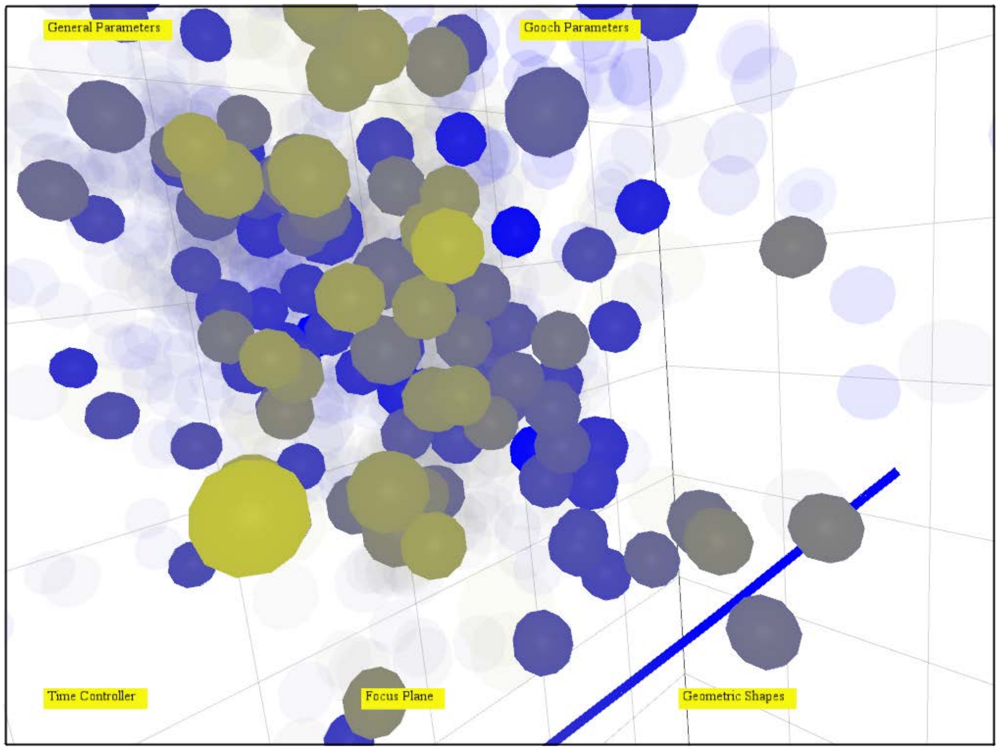

Figure 3.11: 3D visualization showing the events being rendered using extended Gooch shading (yellow-blue relative to their near-to-far distance). The filtered out events are rendered transparent to keep the context.

3.5.4 Early Prototypes

During the early phases of this work, it was not clear in our mind what are the best representations of the data and the best ways to visualize it. Accordingly, we decided to use sketching to try some ideas quickly. Indeed, we implemented some of these ideas as early prototypes (Figure 3.10 shows the old interface of *FractVis*). In this section, we will present some of these early prototypes and discuss them.

Gooch Rendering

We tried to empower users while navigating the 3D microseismic data. We know it is difficult to overcome some 3D problems such as lack of depth perception. Therefore, we tried a different rendering technique that may enhance depth perception. We supported rendering the point events through depth-based Gooch shading [39]. In our implementation, microseismic events are rendered with their colors being modified according to the distance from the camera eye in order to simulate depth perception (Figure 3.11). Our initial evaluation showed that experts did not considered this rendering to be very expressive, so we decided not to consider this rendering further.

Microseismic Geometric Exploration

Early discussions with our domain collaborators highlighted the importance of doing 3D analysis of the microseismic events. Therefore, we tried to provide expert users with a feature to let them explore the 3D connectivity of the events which might give insight about the geometric behavior of the fractures and the population of the microseismic events (Figure 3.12). We implemented 3D-Tree view (Figure 3.10 (4)) acting as an additional visual to show the 3D geometric connectivity for a (sub)set of events. We tried two different connectivity structures (trees/graphs). Balltree (Figure 3.13 (b) and 3.14 left) [74] was our first 3D geometric structure which bounds a subset of event-nodes similarly to the hierarchy of a binary tree structure. The balltree structure enables similarity-like analysis by creating a graph connecting the selected event nodes as defined by the user. For example, visualizing a 3D balltree for any subset of the events might support the identification of similar events. Another geometric structure that has been implemented is 3D Histogram Tree (Figure 3.13 (a) and 3.14 right). Histogram Tree bounds each group of events who share the same *time* value in one sphere, and connects them using solid lines. Each distinct group (reflecting different time-stamp) is connected to the next/previous group using dashed lines. This type of rendering may enable easier identification of event-clusters grouped according to their *time* values. Finally, we supported shifting these renderings from the main 3D view to another view so as not to clutter the 3D view with many visualization features.

Early preliminary evaluation regarding the feature of 3D geometric representations showed that most of the participants were confused about it, and they preferred a simplified correlation in a 2D Scatter plot-like visualization. Accordingly, we decided not to extend this feature further. In addition, one of the participants expressed that we should minimize the cluttering of the GUI by not having many views visible all the time. The same participant highlighted that he expects the

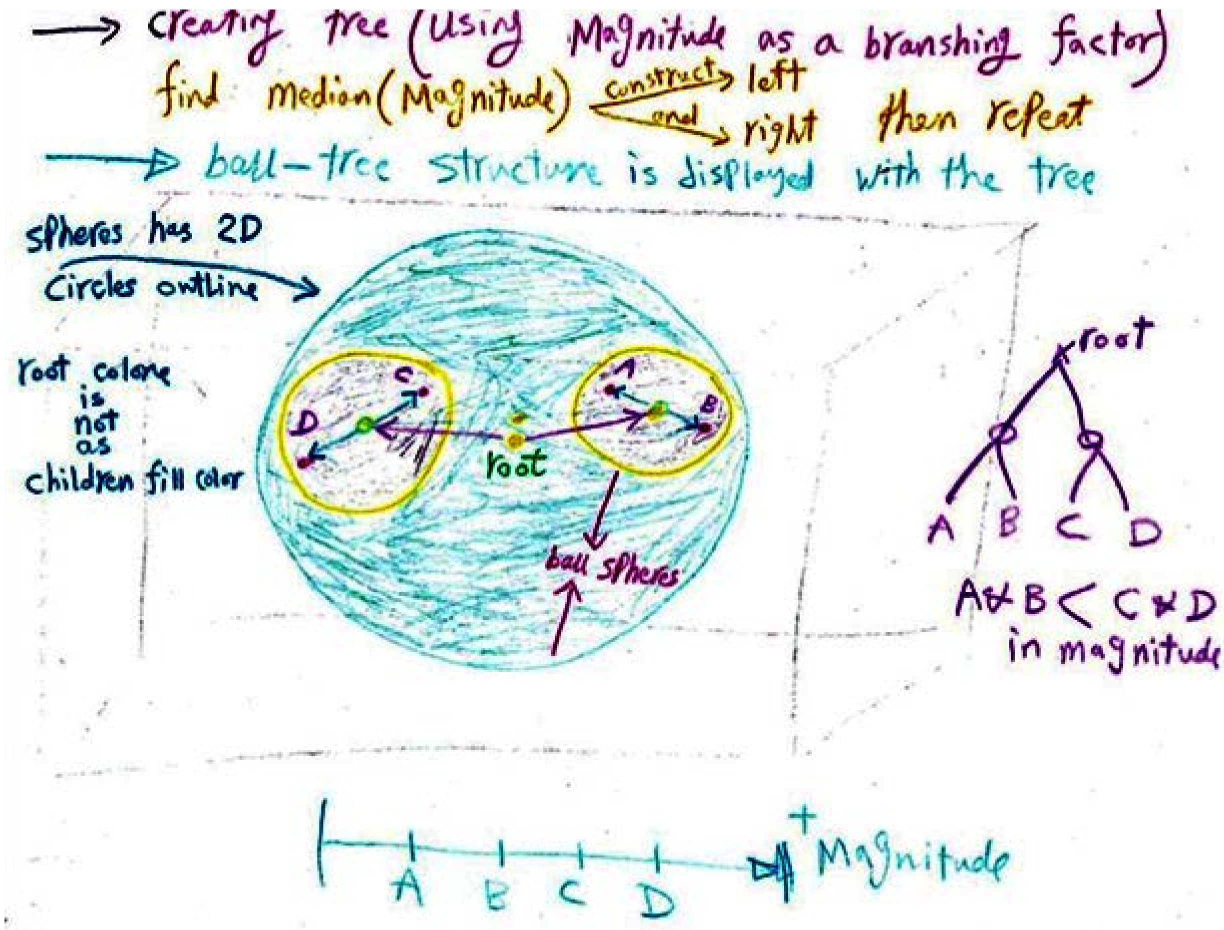

Figure 3.12: A sketch showing the concept of visualizing 3D geometric structures such as a balltree for a subset of the microseismic events

main interface to show only the main visuals (e.g., the 3D visualization and the parallel coordinates visualization).

Region of Interest

According to our collaborator experts, during the analysis there may be a need to work with a subset of 3D events (in a certain region), such as when the expert is considering building the SRV and he would like to ignore the (noisy) unimportant events. Such requirement motivated us to think how to support the experts in selecting events in a region. Indeed, we implemented interactive filtering as one way to refine the data subset.

We also tried the concept of a region of interest (ROI) [89] where further processing will only

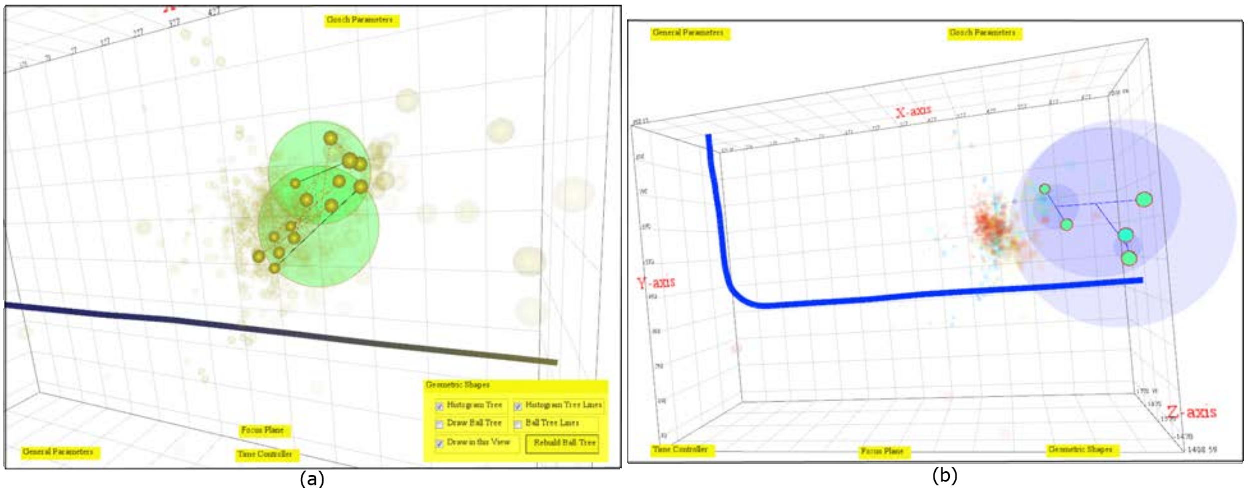

Figure 3.13: Supported geometric representations for a subset of events: (a) 3D Histogram-tree (b) 3D Balltree

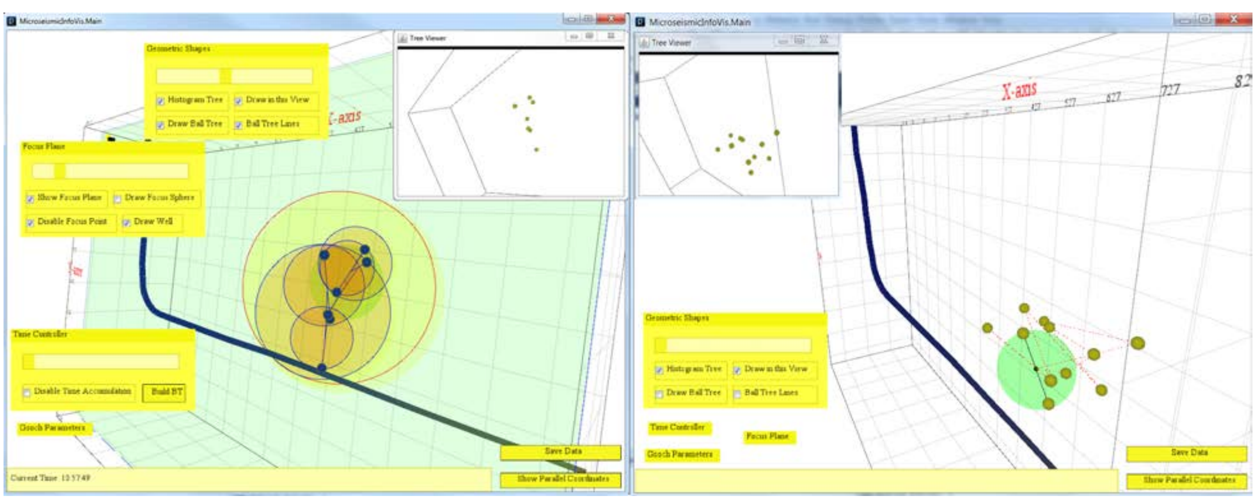

Figure 3.14: Balltree structure (left) and Histogram-tree structure with ROI in light green (right)

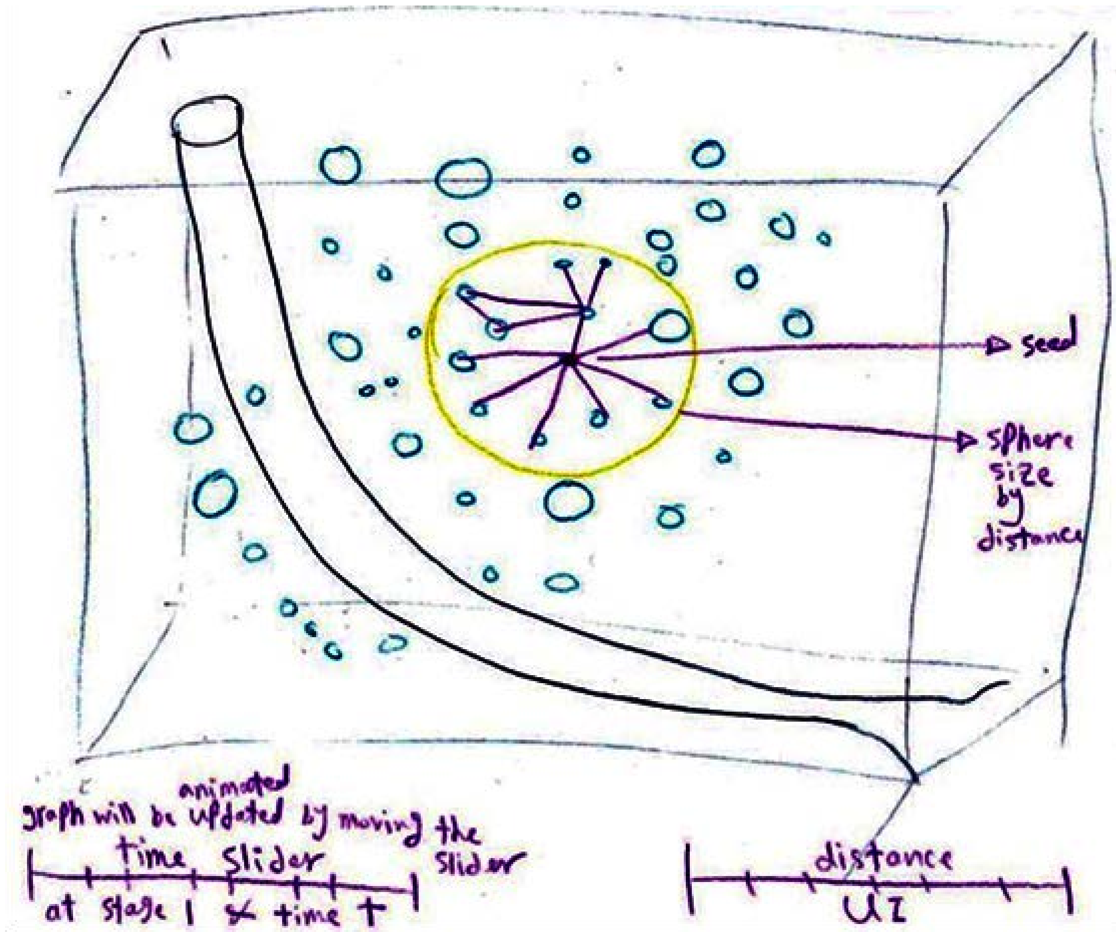

Figure 3.15: A sketch showing the concept of visualizing the microseismic point cloud as spheres where their radius encodes some attribute, and the ROI specify a 3D area where manipulating the events will affects only the events inside that area

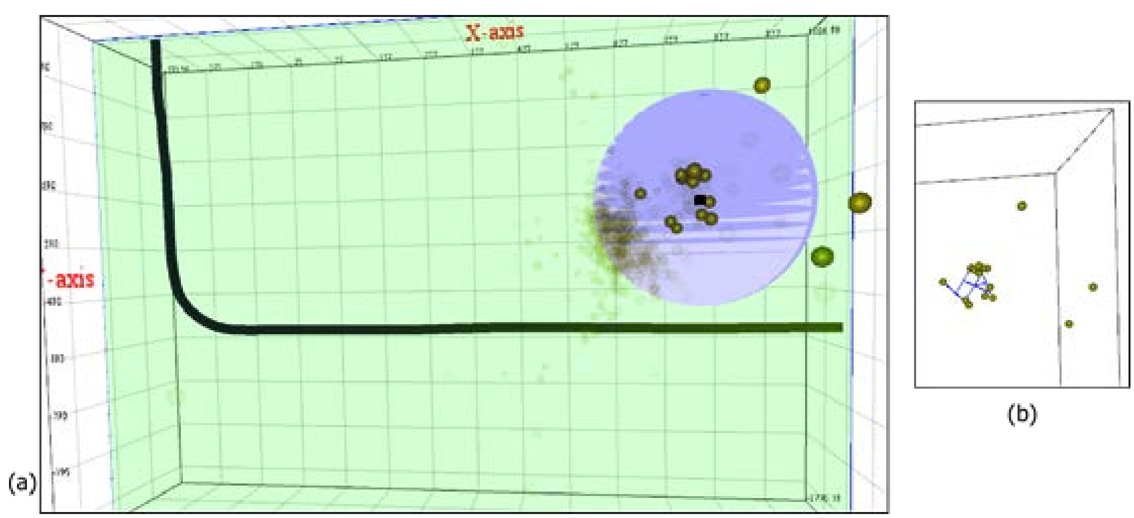

Figure 3.16: (a) Showing the focus plane (green) with a subset of events inside the ROI (violet) with (b) the synchronization of the generated geometric tree in a separate view.

be applied to events inside that area (Figure 3.15). The user can specify a ROI by selecting a focus point over a focus plan, and customizing the region's size/depth. After defining that, he or she can apply some operations which will affect only the events lying inside that 3D area (Figure 3.13 and 3.16 (a)). In one of the early assessment sessions we had, one participant expressed that the way we supported ROI selection is more intuitive than some other (commercial) tools that s/he is using for doing the same task. Although this idea may have a potential for improvements, we did not take it further in our work since we wanted to focus more on the multidimensional correlation aspects.

Tabletop

To improve the user experience, we also implemented a tabletop version of our early prototype (Figure 3.17). The tabletop interface differs from the regular interface because it supports different forms of input. For instance, the expert can interact through touch and/or tangible objects (e.g. tags) to control the visualization. Furthermore, multiple experts can work together and perform collaborative visual analysis of their data leading to improved productivity. We agree that the interaction possibilities available through tabletop present big potential. However, we decided to focus later on multidimensional exploration so we did not consider this approach further.

3.5.5 Implementation

We implemented our visualization system using C# and SharpDX. Rendering thousands of 2D lines in the screen interactively affects the application performance. Therefore, we explored SharpDX as a high performance rendering framework, and used it to render the 3D microseismic events and the 2D visualizations. We also used some external libraries to facilitate building multiple views and managing them in order to improve interaction. Our earlier implementation used Java and the "Processing" library among other external libraries such as "XlsReader" for reading the data files. We decided to port our system because "Processing" did not provide stability and flexibility when working with large 3D data.

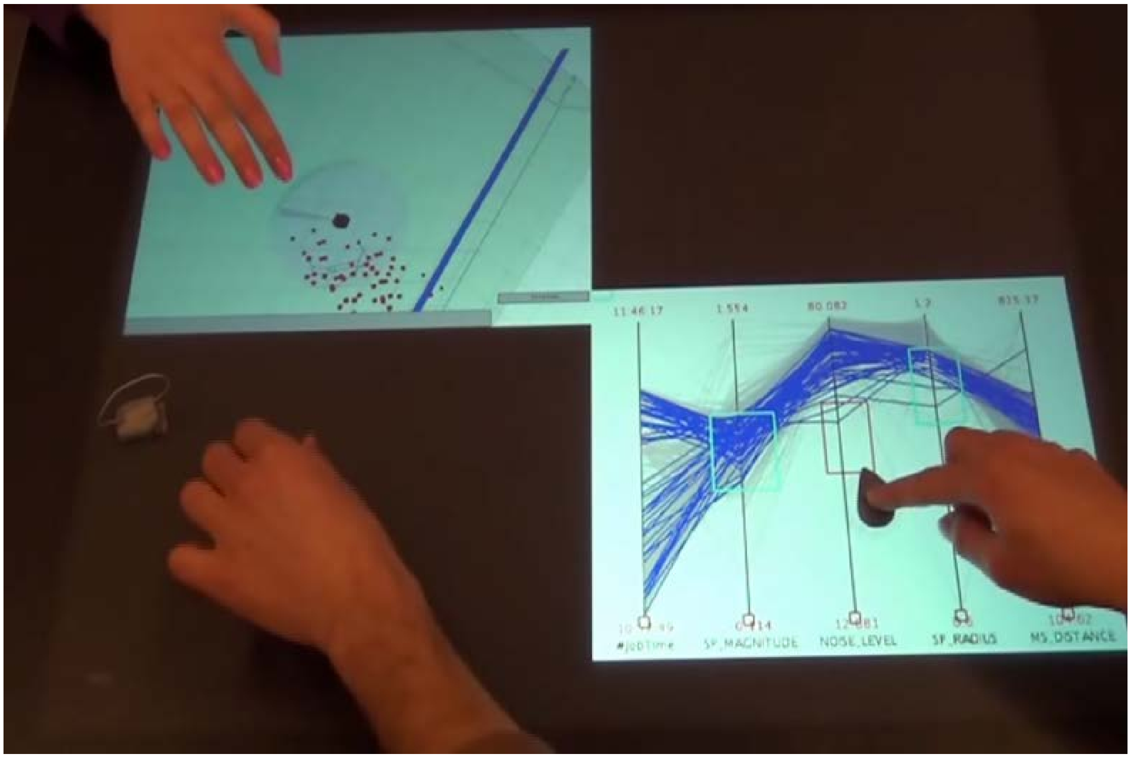

Figure 3.17: Users collaboration over the Microsoft Surface tabletop; one user is filtering the microseismic events while the other is exploring the 3D resulted visualization

Our implementation is flexible and can adapt to any new microseismic data file. Indeed, we visualized another microseismic dataset using our system (Figure 3.18). The experts expressed that our visualization provided them with insight regarding the distribution of the microseismic events in relation to the wells. Their initial findings, using *FractVis*, highlight a systematic difference for the events around each well. They are very excited to continue using *FractVis* and they hope that it would support them further, to visually analyze their data.

3.5.6 Illustrative Example

In this section, we show through an illustrative example (of a typical microseismic scenario) how our prototype can be used to analyze the data. A standard and common analytical task is to create a subset of the data that has minimal outliers. We define any microseismic-event-outlier to be the event with values which are very far from all the other events' values with respect to many attributes (e.g., it can not be clustered). Such a task can be decomposed into key operations. First,

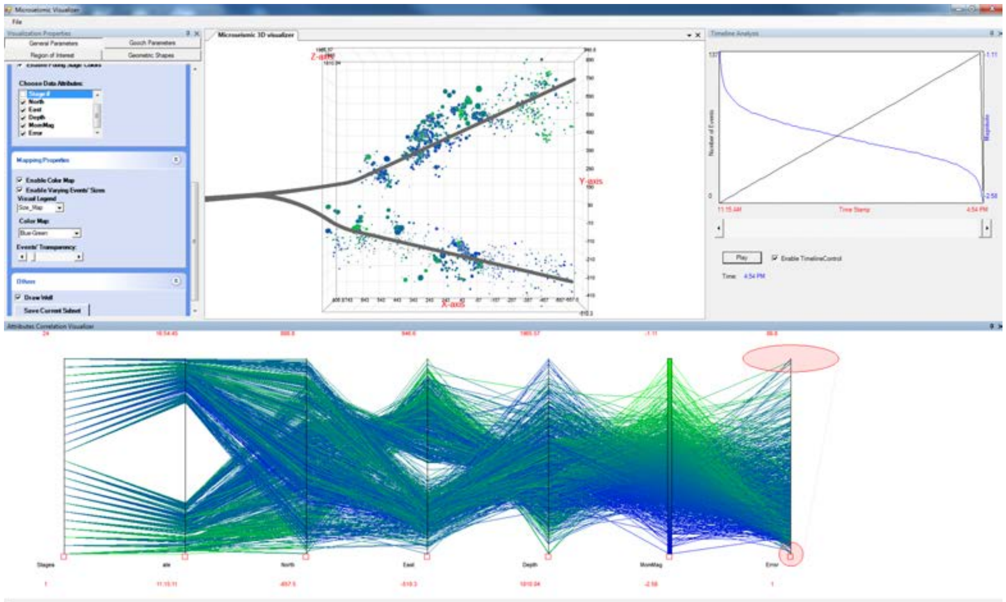

Figure 3.18: Visualizing another microseismic dataset using *FractVis*. The 3D visualization shows that the events from well A (top) are systematically higher than those from well B (bottom).

we need to find the possible outliers or anomalies. Second, we need to perform a test to confirm that they are outliers and not just some uncommon values. To achieve this operation, we may need to correlate some of the data attributes. Third, we need to filter out the confirmed outliers and save the refined subset. Following we describe this illustrative example.

After starting our tool, we will notice that there are two main visualizations, 3D events' visualization and the PC visualization of the microseismic data attributes. First, we examine the PC visualization and we may notice some possible polylines who are not part of the data clusters. These could be possible outliers. Second, we need to check and confirm each possible outlier. We repeatedly create a filter-box that shows only each possible outlier by showing only its polyline (line-path). Then we examine the path of the polyline to see if it is an outlier over other attributes or not (Figure 3.19 left). If yes, then it is confirmed as an outlier. This process needs to be repeated for each possible outlier. If we are not sure about some events' line-paths, then we can use the feature of "color mapping" or "size mapping" to further analyze the events geometry and correlation. For example, applying the feature of "size mapping" will allow the user to inspect the 3D event location to determine from the geometry if it is an outlier or not (i.e., the location of the event is

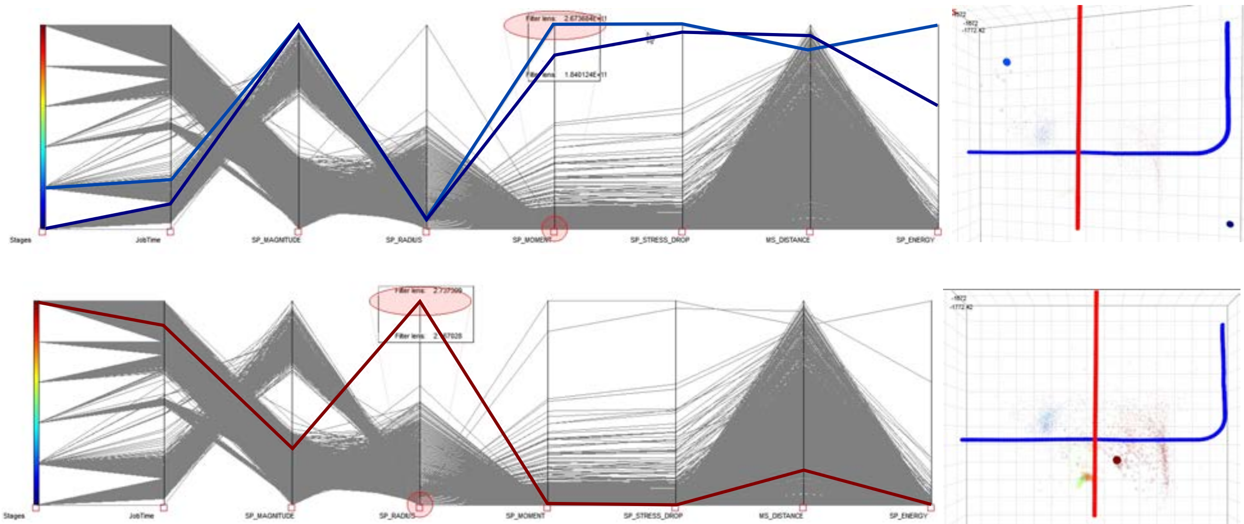

Figure 3.19: Illustrative examples about outlier analysis. Top: two events (blue line paths) are confirmed as outliers. Botton: one event classified as non-outlier although it has uncommon value since it is near the monitoring well

very far away from the treatment well, see Figure 3.19 right). Finally, we create filter boxes to select all the events except the confirmed outliers, and then we save the data subset.

3.6 Evaluation

Following the methodology described in [83], we had early consultation sessions with domain experts before starting this project and we identified some of their practices and needs. Then, we followed an iterative (incremental) design approach (with continuous ad hoc evaluation) while developing our system to insure that it meets their expectations. This methodology, in our opinion, resembles a self-validation approach. We conducted preliminary evaluation by demoing our visualization prototype to the domain experts and also to visualization researchers. In this section, we describe the study procedure (Section 3.6.2), the participants (Section 3.6.1) and detail some of the results that we gathered and analyzed (Section 3.6.3).

3.6.1 Participants

We had six participants; two of them are visualization researchers and four of them are domain experts in which two of them are highly experienced with 5+ years in the microseismic industry (domain collaborators). One of the highly domain experienced participants is internal to our group and he or she provided us with continuous valuable feedback and comments throughout the development cycle. The two visualization researchers are computer science graduate students who are familiar with visualization and graphics techniques. The reason that we had visualization participants is that we wanted to get some feedback about the novel visualization features in our system.

3.6.2 Study Procedure

We conducted eleven assessment sessions to gather (and analyze) the feedback of the participants including their reaction regarding *FractVis*. We did not have formal sessions but rather a couple of assessment sessions similar to focus group meetings and design critique sessions. The duration of each session was from 60 to 90 minutes. At the beginning, we had open-ended free form design critique sessions. Then, by the middle of the development, we added a pre-session questionnaire in order to check & validate our understanding about the needs of our collaborators. After we finished the development, we incorporated post-session questionnaires (through a semi-structured interview) in our assessment sessions in order to gather explicit feedback about *FractVis*' features (Appendix B contains more details about the evaluation study including the used questionnaire). The reason for having different evaluation types is that we tried to adapt the inquiries of each session to the level-of-information that we would like to attain. In addition, some evaluation session have been repeated with some of the participants (in later times) for two reasons: (1) it is difficult and challenging to get many (different) domain experts available each time you want to evaluate *FractVis*' features, and (2) we wanted to validate that our continuous improvements of the system addressed the issues raised by some of the participants during the early sessions we had with them.

After having our first prototype, each session included a demonstration about our tool, and then we gathered the participants feedback and reaction. We video and audio taped the sessions (with participants consent) to help us later in the analysis. We transcribed the sessions, identified key verbal comments, and analyzed participants' responses to highlight strengthens, weaknesses and any suggestions for improvements.

3.6.3 Results

We demonstrated our tool in different sessions, and we analyzed the reaction and the feedback of the participants regarding our tool. In this part, we describe the feedback regarding our approach strengths, limitations and any directions for improvements. We also highlight some of the participants' noteworthy comments as well as general feedback shared by most of them.

Most of our participants provided positive feedback about many of *FractVis* features. One of the highly experienced domain experts discovered a very interesting issue with the data calculation using our visualization. The expert analyzed the relation: *magnitude* vs. *distance*, and specifically expressed:

"When I look at this, I can see there is a problem with the data ... because it is not physically feasible ... So this just highlights some problem with the data".

The expert also highlighted that our features are not available commercially (as per the domain concerns), and commented about that saying:

"I had different experiences with different commercial packages... I think what you are doing here is quite unique and different; not available commercially".

At the beginning, one of our domain collaborators complained that the GUI of *FractVis* is cluttered and it needs to be simplified. After updating *FractVis*, we had another assessment session with the same domain collaborator. The collaborator commented saying that our tool has GUI improvements and it is not cluttered anymore. This may reflect that our continuous improvements led to satisfying the expectations of the experts. A third (domain) participant expressed that the feature of "shadow boxes" is good for keeping the context of the nearest events according to their *time* attribute. We asked all participants to rank the prototype features and the usefulness of using *FractVis* for microseismic analysis in general. Four out of six participants filled the questionnaire and all of them *"strongly agreed"* that the dynamic-boxes' feature is useful. Most of them also *"strongly liked"* the PC, while one participant *"slightly liked"* it. Furthermore, all of our participants *"strongly liked"* the feature of color-correlation and *"slightly liked"* the feature of size-correlation. Finally, we asked them how our tool can be useful for the analysis of the microseismic data. Half of them said it is strongly useful while the remaining said it is slightly useful.

Another feedback that shows some limitations and weaknesses in our prototype has been provided by some of the participants as well. One domain expert expressed her opinion about our feature of having embedded visuals within the PC as confusing. The expert specifically expressed:

"I like it popped up in the middle, but what it did is just disconnected the way I am looking into the data so I have to go back".

The participant also expressed that the data units are missing and it is expected to have the proper units (i.e. metric) rendered beside each data value for clarification. Another domain expert highlighted some limitations in our system regarding the ability to perform free-form interaction with the 3D events.

We also received lots of suggestions for future improvements of *FractVis*. When one of the domain experts saw our scatter plot visualization within the PC, s/he noticed that the colors of all the dots are black while the PC's polylines are colored differently. S/he suggested that the scatter-plot's dots to be colored matching the PC's polylines in order to make the coordination between them easier. Our current implementation of the filter boxes considers only the intersection (ANDing logically) of the results, but a suggestion came here to enable supporting different logical operators between the multiple filter boxes. Another domain expert expressed that is will be very useful to integrate some automated statistical features into the visualization of *FractVis*. S/he also suggested that integrating additional types of data (i.e. engineering curves) would be important. A similar suggestion came from another domain expert to improve the time-based visualization, with functionality such as playing-back the time & controlling the events' population, because this can be very handy.

To summarize, most of our participants and domain collaborators expressed that *FractVis* is useful and our post-session questionnaire clearly highlights this feedback.

3.7 Discussion and Lessons Learned

In this section, we present the main insight and lessons we have learned during the design and evaluation of *FractVis*. We also reflect on the difficulties that we have faced, and then conclude with set of guidelines and suggestions for future development of similar problem-driven visualization systems.

Designing and implementing problem-driven visualization systems such as *FractVis* can be challenging due to incorrect understanding of domain's requirements and needs. We stress here how crucial it is to conduct thorough contextual inquiry [44, 77] sessions with domain experts before commencing the system design and implementation. We followed the methodology described in [83], and we agree that it is essential to start the project by attempting to understand the domain processes, requirements and challenges.

Given the relatively recent emergence of microseismic monitoring methods, the number of domain experts is limited. Clearly, having access to a limited number of domain experts may not be suitable for conducting detailed formal evaluation, but it does suggest other benefits. First, it is easier to contact and coordinate with such small group of experts. Second, the repeating sessions with the same experts allowed for continuous and coherent feedback and refinement of the prototype. Third, having repeated access to the same experts allowed us to confirm that the system features meet their expectations. On the other hand, we recommend the following strategies as general practices to simplify the issue of having only few experts. We argue that it would by helpful to integrate the system in the working environment of the available domain experts and to collect continuously their feedback. Doing so would give the domain experts the chance to work with the system for some time, to familiarize themselves with it, and to be prepared for providing more insightful feedback later. Furthermore, conducting repeated sessions with the same experts during the different project phases is useful for providing continuous feedback. We consider this as a self-validation approach.

Generally, throughout the process we felt that domain experts are resisting considering and learning new tools as new ways of analyzing their data. While we understood this reaction, it was one of the main challenges that we have faced. Indeed, it inspired us to think about simplifying our design in order to provide experts with a simpler, yet still familiar, tool with the hope of supporting them gaining new insight. One example of such experience was when we introduced PC as new visualization to them. Our experts were not familiar with PC, and they seem to resist understanding or using it in our early sessions with them. Following this initial resistance we provided the experts with additional visualizations which were more familiar to them, such as scatterplot, integrated with the PC visualization. Our approach was that embedding the new visualization side-by-side with familiar ones would allow users to explore it while retaining a known baseline context, allowing them to learn the new technique. The feedback that we received (from most of the participants) confirmed that our approach was useful and helpful. Overall, we wanted to empower the PC visualization by adding the flexibility to see additional (embedded) visuals which would lead to enhancing the data analysis experience.

During some of the assessment sessions with the domain experts, they commented about having many different visualizations and interaction possibilities in our system. Some of them considered that to be confusing and they just preferred simple visualizations, while others considered it to be a form of flexibility. One of the comments regarding PC is:

"The parallel coordinates is very unique, and you've just showed me that it can be more powerful... when I become a good user with it, by grabbing the axis around to see what I want, and crossplot everything and so on, it will be tremendously useful".

On the other hand, when we asked a participant about the idea of having multiple dynamic (filter) boxes, and whether it is easier or not, s/he specifically replied saying

"That would be something that I have to use for some time to know if it would be easier or not, but for now I think the concept is useful. I think it can be a very good idea".

These comments suggest that our prototype may be a good start for microseismic visual-analysis, though a detailed and formal evaluation is needed to fully confirm that and guide future development.

Finally, one of the highly experienced domain experts discovered through *FractVis* some problem with the dataset calculation: the existence of two *magnitude* clouds in the data, while only one is expected. This finding represents the type of easily accessible insight that *FractVis* could provide, and we believe that further usage of our system especially in the domain-environment would enrich the microseismic data analysis process.

3.8 Future Work

Since it is an ongoing project, there are many improvements to follow. As future work, we are considering the suggestions of the feedback received, regarding improving the prototype and adding additional important features. For instance, we plan to extend the feature of dynamic boxes to support the experts with the ability to manually control the range of each dynamic-box. We are planning also to add more types of dynamic boxes to support showing statistical features within the visualization. Regarding the 3D interaction, we plan to add the capability of screen-sketching with the hope of improving the interaction with the 3D events. Furthermore, it would be useful to enable the user to directly interact with the individual events and be able to see details-on-demand. Extending the time-based visualization by integrating engineering curves and synchronizing them with the other visuals of *FractVis* is also one theme of our future work. Finally, we plan to conduct a formal detailed user study to obtain a more comprehensive feedback regarding the user experience with our prototype. This would also include conducting ethnographic sessions with the microseismic domain experts to refine our understanding of their processes and practices. In essence, our ultimate goal is to create a complete visual-analytic solution for the microseismic experts by extending and integrating other visualization techniques.

3.9 Summary

In this chapter, we present our efforts to support the visual exploration of the high dimensional microseismic data. We detail a characterization of the microseismic monitoring domain including data and task abstractions. Based on that, we also explain a set of design requirements and visual representation choices specific to the development of microseismic visualization. Furthermore, we describe the tool we developed, *FractVis*, to support the visual analysis of the microseismic data. Our tool is composed of a set of coordinated visualizations that resulted by combining and extending different techniques through an iterative collaborative process with the experts. We adapt the technique of parallel coordinates through proposing two novel features in order to enable intuitive data filtering and attributes' correlation. Although our tool represents the first prototype and there are lots of improvements to follow, our preliminary evaluation showed that insights could be gained from it. Beyond the developed characterization and visualizations, we show through an illustrative example how our prototype can be useful for analyzing the microseismic data. We also highlight from our experiences some of the lessons learned and guidelines that could be useful for extending this research. Provided ideas and suggestions also motivated several instances for future work. And so, the upcoming chapter reports on follow up work on microseismic visualization, focusing on exploring new forms of interaction to improve the visual analysis of the microseismic data, by applying recent interaction mechanisms including the use of proxemics and a painting metaphor to navigate and manipulate the 3D events.

Chapter 4

Proxemic FractVis: Interacting with Microseismic Visualization

Microseismic visualization systems present complex 3D data of small seismic events within oil/gas reservoirs to allow experts to explore and interact with that data. Yet existing systems suffer several problems: 3D spatial navigation and orientation is difficult, and selecting 3D data is challenging due to the problems of occlusion and lack of depth perception. In this chapter, we present our efforts to mitigate these problems by applying both proxemic interactions and a spatial input device to simplify how experts navigate through the visualization, and a painting metaphor to simplify how they select that information.

4.1 Overview

Microseismic experts consider analyzing geological fracture geometry as essential task in their work. However, performing this task efficiently requires them to have an intuitive way to navigate, explore, and select subsets of the complex 3D microseismic data set. Existing microseismic visualization systems typically portray data as a 3D point cloud. Yet navigation and orienting oneself around this data is awkward using traditional interaction techniques since they are limited in providing 3D immersion-like virtual experience. Furthermore, selecting such data in 3D is difficult due to problems such as occlusion and lack of depth perception. Our goal is to improve upon these forms of interaction and support experts with intuitive 3D interaction mechanisms (Figure 4.1).

In the following sections we present our initial efforts in achieving this goal, where we apply proxemic interactions [15] and a spatial input device along with a painting metaphor to ease basic navigation and selection tasks. We also highlight some of the lessons learned and likely improvements.

Figure 4.1: Interacting with microseismic 3D data.

4.2 Microseismic 3D and FractVis

FractVis (Chapter 3) is an experimental 3D visualization system, built to support how microseismic domain experts can geometrically analyze their 3D data. We used its microseismic domain as our context to investigate 3D problems in that domain and how to improve 3D interactions within it. In particular, we identified several important tasks that involve 3D-related issues. One of these is the calculation of stimulated reservoir volume (SRV), which is the volume of rock affected by the seismic stimulation [12]. To perform this calculation, domain experts navigate the 3D geometry of the data, where their tasks include things such as looking for and analyzing the locations of the microseismic events in relation to the well-bores in the reservoir. This includes selecting subsets of that data of particular interest, where they filter out some of these events and extract a 3D subset that will later represent the estimated oil volume.

Performing such calculations, however, requires the domain expert's ability to interact through the complex GUI of the microseismic visualization system. For example, a domain expert has to

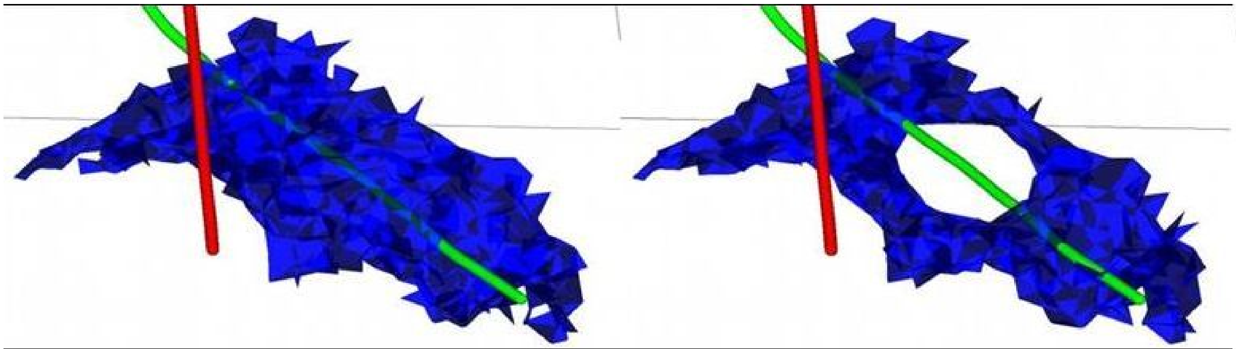

Figure 4.2: Excluding some part of the SRV through sketching followed by extrusion [12].

explore the 3D space using the mouse along with many keyboard buttons and GUI combinations in order to sketch a 2D area. The sketched 2D area is then extruded with full depth, to generate a volume, in order to select subset of the data (Figure 4.2). Although this approach is being used now, it is awkward and requires considerable training. Furthermore, it has many limitations regarding data selection. For instance, the experts cannot control the depth level of selected area.

Our approach considered the design of *Proxemic FractVis*, a prototype which is only roughly similar to a CAVE [29]. We used existing (widely available) technology to visualize the data, a Wiimote controller, and motion capture sensors (Vicon). We used these technologies in two ways: first, we applied proxemic interactions to mediate the interaction with the 3D microseismic data, ultimately to make it more natural to explore and navigate around the 3D data. We tracked many proxemic dimensions such as distance, location and orientation between the person and display to facilitate the interaction. Similarly, our painting metaphor attempts to ease selection of subset of the data up to a specific depth level. The simple idea is to let the user navigate through the virtual content of the screen by relating his distance and perspective and update the visualization accordingly. Second, we used the Wiimote controller as a device to fine-tune the navigation and interaction with the data by 'painting' it (Figures 4.3). While others have used Wiimote for painting (e.g., [54]), we use it to let people select the 3D microseismic data. In essence, our approach combined different interaction techniques to simplify and enhance the 3D interaction for microseismic domain.

Figure 4.3: Physically based virtual painting [54].

We decided to illustrate our interactions through *FractVis* for many reasons. First, *FractVis* uses 3D data and suffers from the common 3D problems (e.g. occlusion and lack of depth perception). Second, the expected users of the system are experts in the oil and gas domain who usually have access to visualization rooms (such as CAVE) and virtual reality technologies (e.g. [103]). The availability of such an environment will make it easier to evaluate the developed prototype. Accordingly, we extended *FractVis* to showcase our new interactions as explained below. However, we believe that our approach can be generalized to other 3D visualizations that support navigation and data selection.

4.3 Implementation

Proxemic FractVis uses the Vicon hardware to track and capture user's movement and interaction. We also used Wii as our pointing tool (Figure 4.4) and we associated some markers with it to allow real time capturing and tracking (Figure 4.5). The tracked information is preprocessed and is available to us through the Proximity Toolkit [60]. However, the data received from the proximity toolkit sometimes has noise associated with it. Therefore, we implemented a smoothing mechanism to rectify the input and to provide our system with a stable noise-free input.

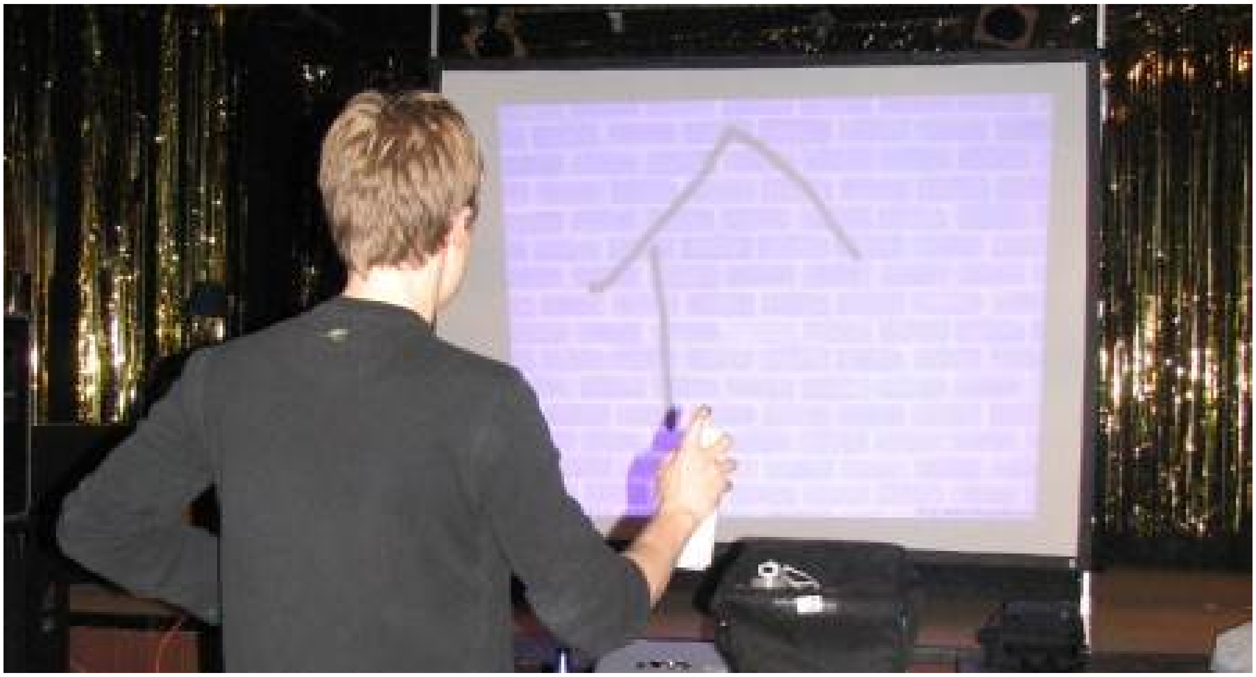

Figure 4.4: Virtual painting using Wii mote [7].

Figure 4.5: The Wii controller (with attached markers) used in our prototype for pointing and interaction.

4.4 Navigation

4.4.1 Coarse Navigation by Proxemics

Our approach immerses the expert inside the *FractVis* 3D world, where the expert can navigate around the 3D data. That is, we map the 3D scene to the bounds of the room, and we transform the scene as a function of proxemics, i.e., the expert's distance, location and orientation relative to the display. The 3D visualization is continuously updated relative to its proxemics relation to the expert. For instance, the distance between the expert and the vertical display is used to control the level of detail of the visualization. That means, when the user is near, the scene is zoomed-in to provide more details and when the user is far the scene is zoomed-out to provide fewer details. The camera responds to the location and orientation of the person relative to display by rotating the scene so that its 3D content always align with the expert's view of it.

Figures 4.6 and 4.7 illustrate this basic navigational. In Figure 4.6 (bottom), the expert is approaching the data volume, where he sees it in its entirety. In Figure 4.6 (top), the expert has moved closer to the screen, and the data has smoothly zoomed in to match his approach, thus showing increasing detail. In Figure 4.7 (bottom), the expert moves from to the side to view the data from a different perspective; the scene transforms itself to follow this new viewing orientation.

4.4.2 Fine Navigation by a Device

We observed that tracking the data with a person's body is good for coarse-grained navigation (e.g., for broad exploration of overview, detail, and vantage points) but not for fine-grained navigation. At any time, the expert can 'freeze' the 3D world by pressing a button on his hand-held spatially-tracked Wiimote. The Wiimote then acts as a 3D mouse, where (depending on the button pressed) the now-stationary expert can fine-tune their zoom level and the camera orientation of the data by moving the mouse in 3-space. For example, in Figure 4.7 (top) the expert is moving Wiimote to navigate around the data and see it from different orientations while standing in a specific location. The expert can thus continue to navigate the scene with the Wiimote. In brief, the mental model is

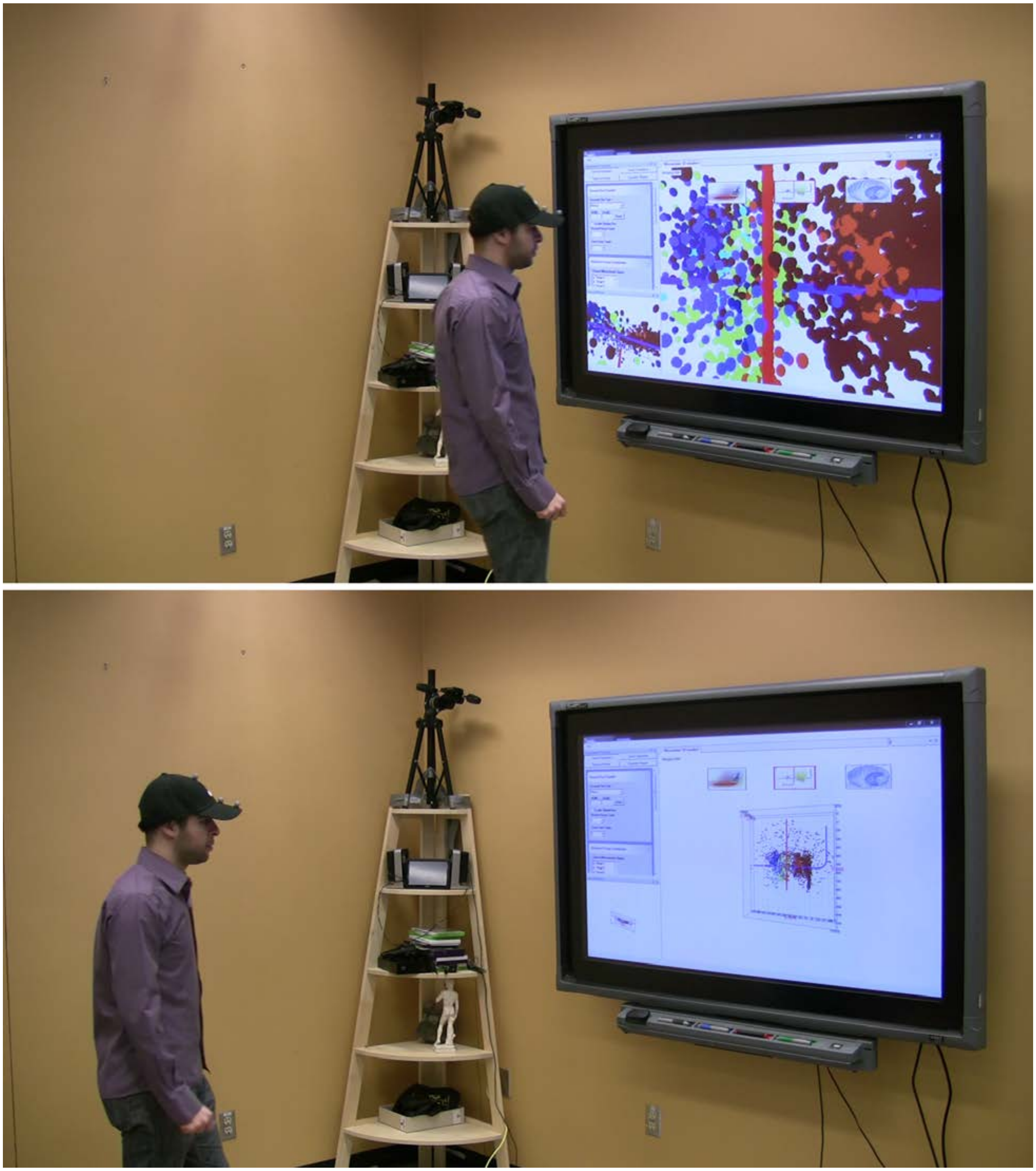

Figure 4.6: Zooming in on approach (top), zoomed out from afar (bottom).

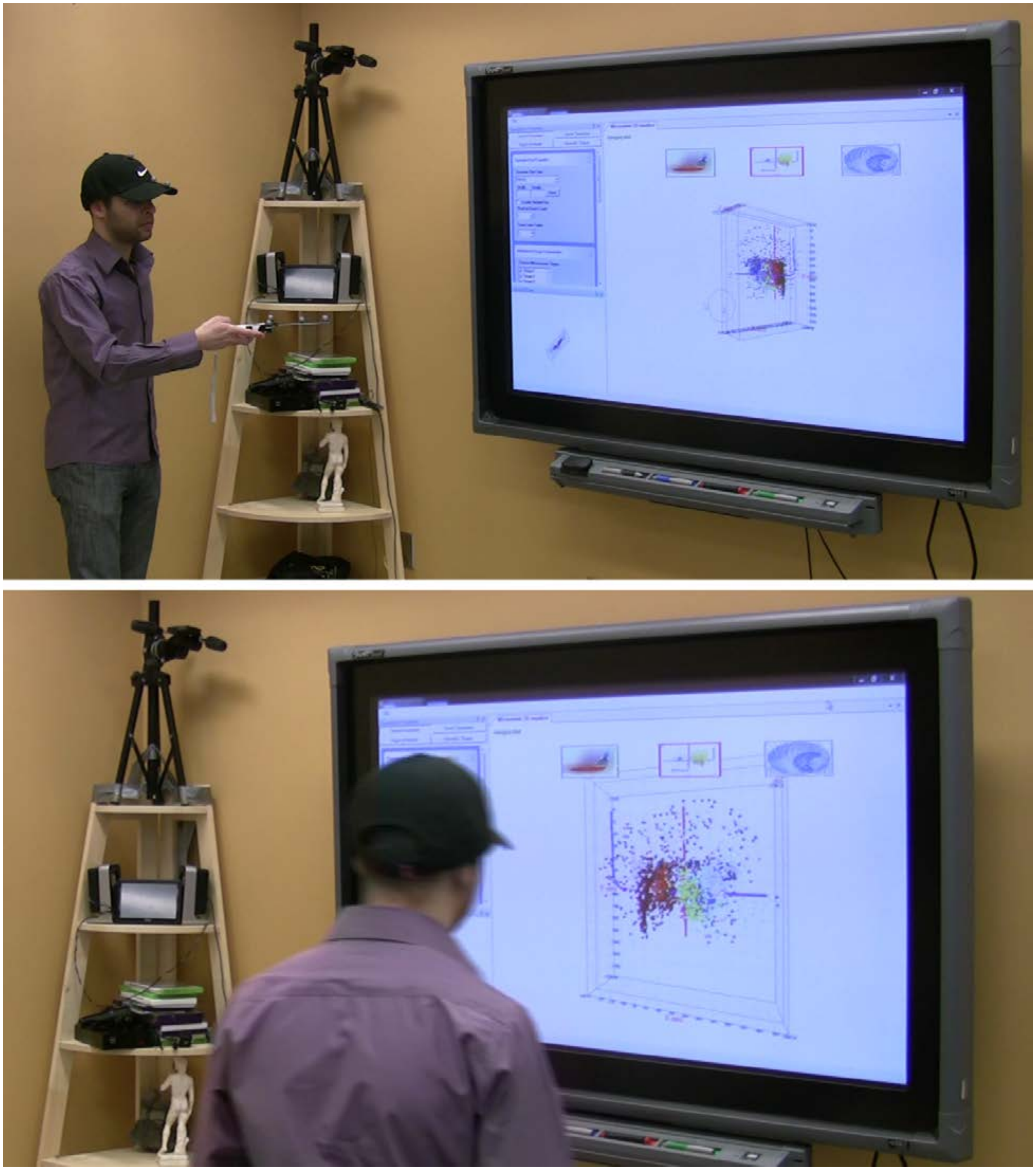

Figure 4.7: Use the Wiimote to fine-tune the data navigation (top), moving to change perspective (bottom).

that the proxemics of the user's body provides coarse navigation, while the Wiimote extends one's hand to provide refined navigation as needed.

4.5 Interaction

4.5.1 Sphere Brushing to Select Data

Our system supports navigation and interaction by mapping the expert's location directly in the virtual world. The mental model here assumes that the expert is imagining oneself inside the 3D world and his location is mapped inside that world as sphere, representing the ROI.

To interact with the 3D microseismic data, the user would start by moving around in his/her physical space while observing his/her corresponding sphere location inside the 3D data cube. After being satisfied with his/her location (which reflects the ROI), he/she can use the Wii controller to initiate actions and manipulate the 3D content (Figure 4.8). For instance, the user could use the Wii to point and define the height and/or the size (radius) of the ROI sphere (Figure 4.9). Then he/she may issue the brushing action to make the system filter the data and show only the events inside the corresponding ROI (Figure 4.8 bottom).

4.5.2 Spray Painting to Select Data

Our system also allows an expert to interact with the data, where he uses the Wiimote to point at particular data and to select it. In particular, the expert can brush the 3D data in order to select it via a spray painting metaphor.

The mental model is that the data exists inside a 3D bounding cube, where painting surface resides inside that cube at specific depth as a rectangular slice (plane). To begin, the expert navigates to the appropriate viewpoint, as described above (Figure 4.7). The expert then uses a different button on the Wiimote to navigate to the desired painting depth, by progressively moving through slices within the cube (Figure 4.10). In Figure 4.11 (top), the expert has oriented himself within the cube, and he begins spray painting (using a different button) to select the desired data. Figure 4.11

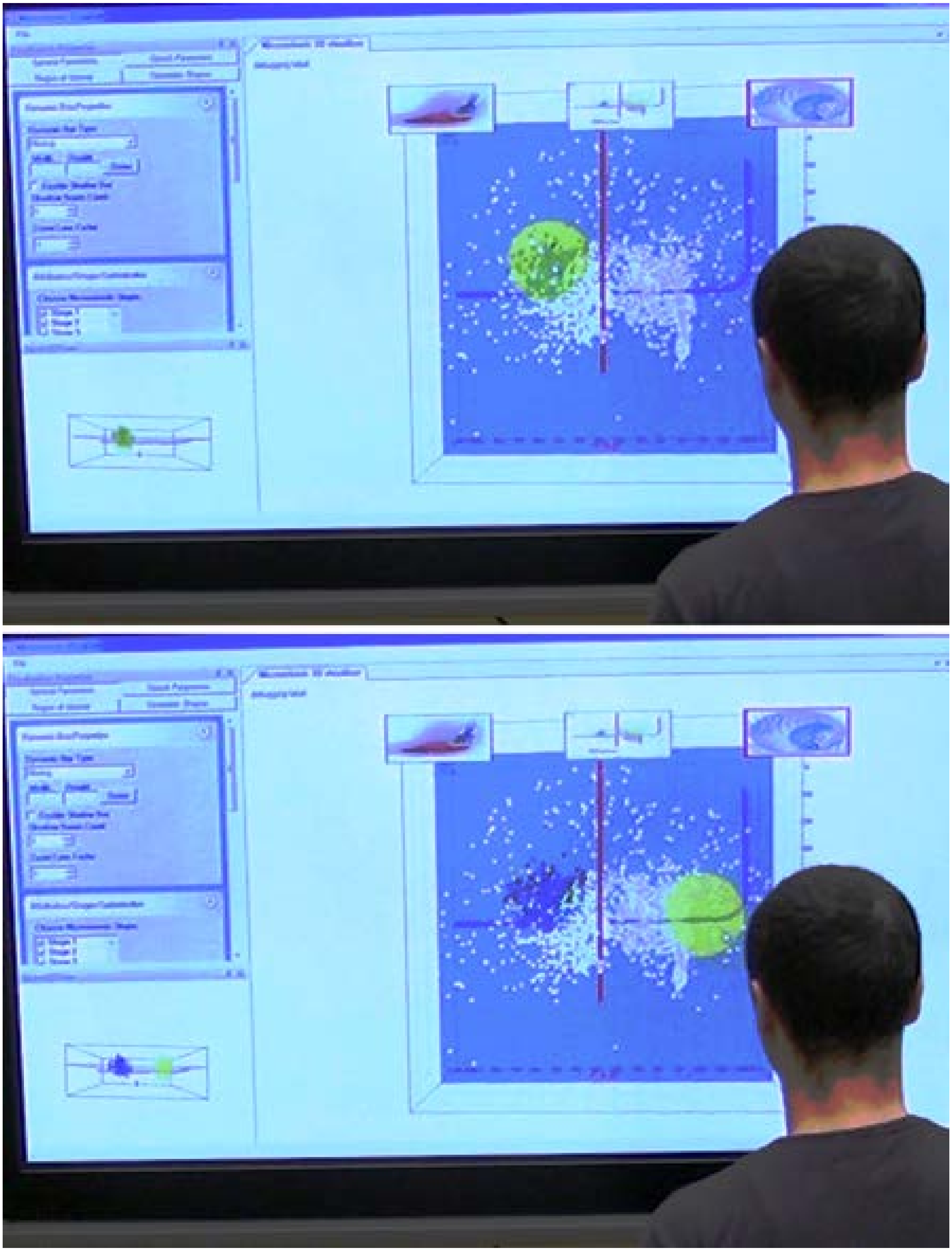

Figure 4.8: Performing sphere-based brushing: (top) the user is about to perform the action, (bottom) the results of selecting only the events inside the ROI after performing the brushing.

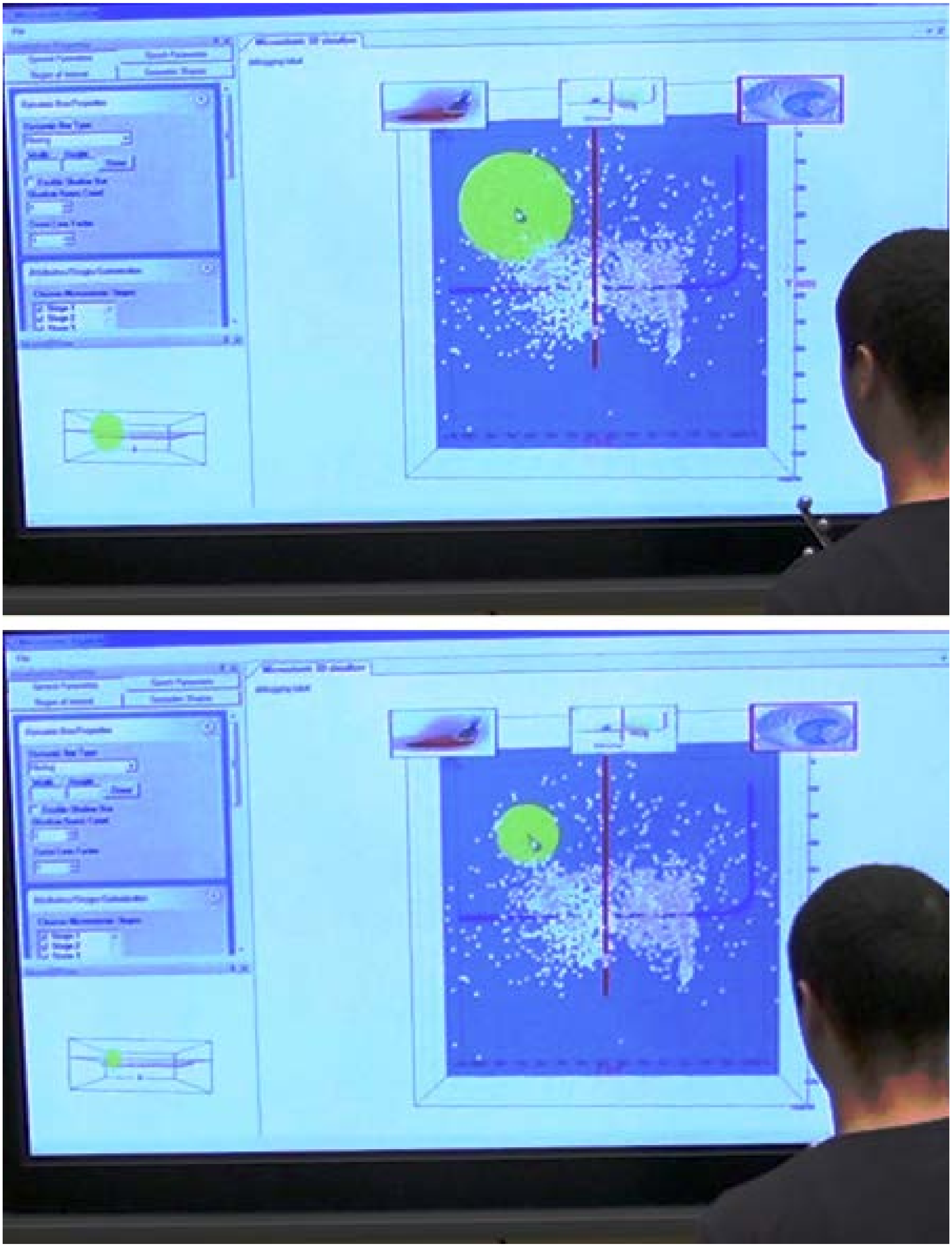

Figure 4.9: Adjusting the brushing sphere's radius (big: top, small: bottom).

(bottom) shows the results, where the selected data is being shown colored. The expert can then continue this process to fine-tune the subset of the selected data (Figure 4.12). Although spray painting is happening over the fixed 2D slice, we use a projection technique to affect the data that exists in front of the painting surface and ignore all data behind it.

4.6 Discussion

Designing and developing new form of interaction by combining different technologies and interaction techniques is difficult. In this section, we outline some of the lessons that we have learned, some of the challenges we faced, and we conclude with our implications to guide future research about improving 3D interactions.

4.6.1 Questions about user acceptance

Our system is a working proof of concept, and as such is not yet ready for a user study. Of course, we believe such a study is required to evaluate and find out more about the practicality of our approach. We expect that our new form of interactions will be resisted by experts who are trained to currently perform this task using a traditional desktop and mouse. We do not expect that our microseismic domain expert will immediately accept the need to stand and move around in order to interact with the 3D data. As usual in these cases, benefits will likely occur only after an expert has gone beyond the initial learning curve, and only when he reaches a level of proficiency that pushes him past what he can do with his traditional desktop-based solutions. Clearly, some form of participatory design will be required, both to elicit the design nuances that domain experts would like, and to develop champions within the community.

4.6.2 Hardware

Our prototype currently uses the Vicon hardware for object tracking. While highly accurate and appropriate for prototype development, the Vicon is quite expensive and as such impractical for

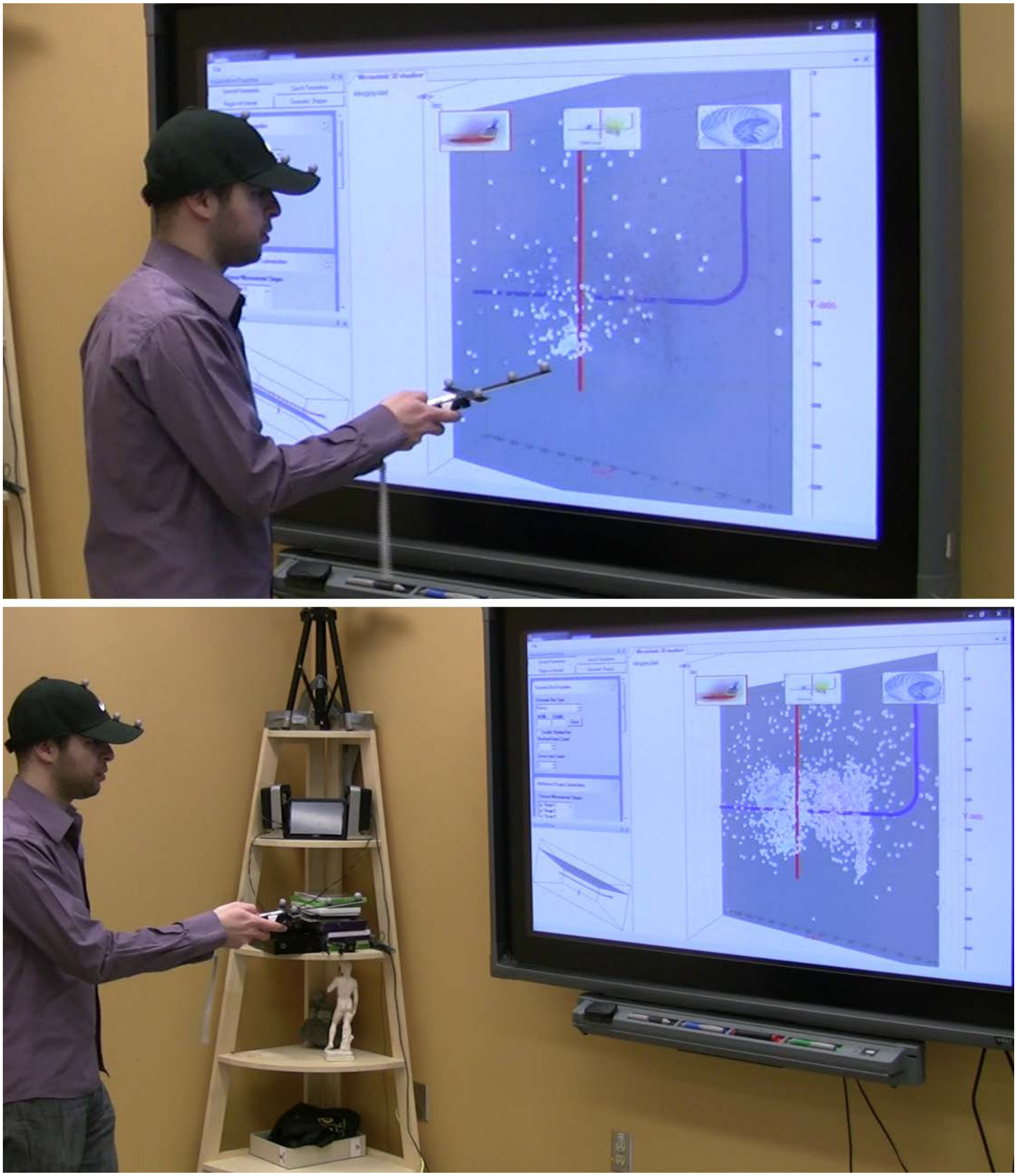

Figure 4.10: Adjusting the painting surface's depth (near: top, far: bottom).

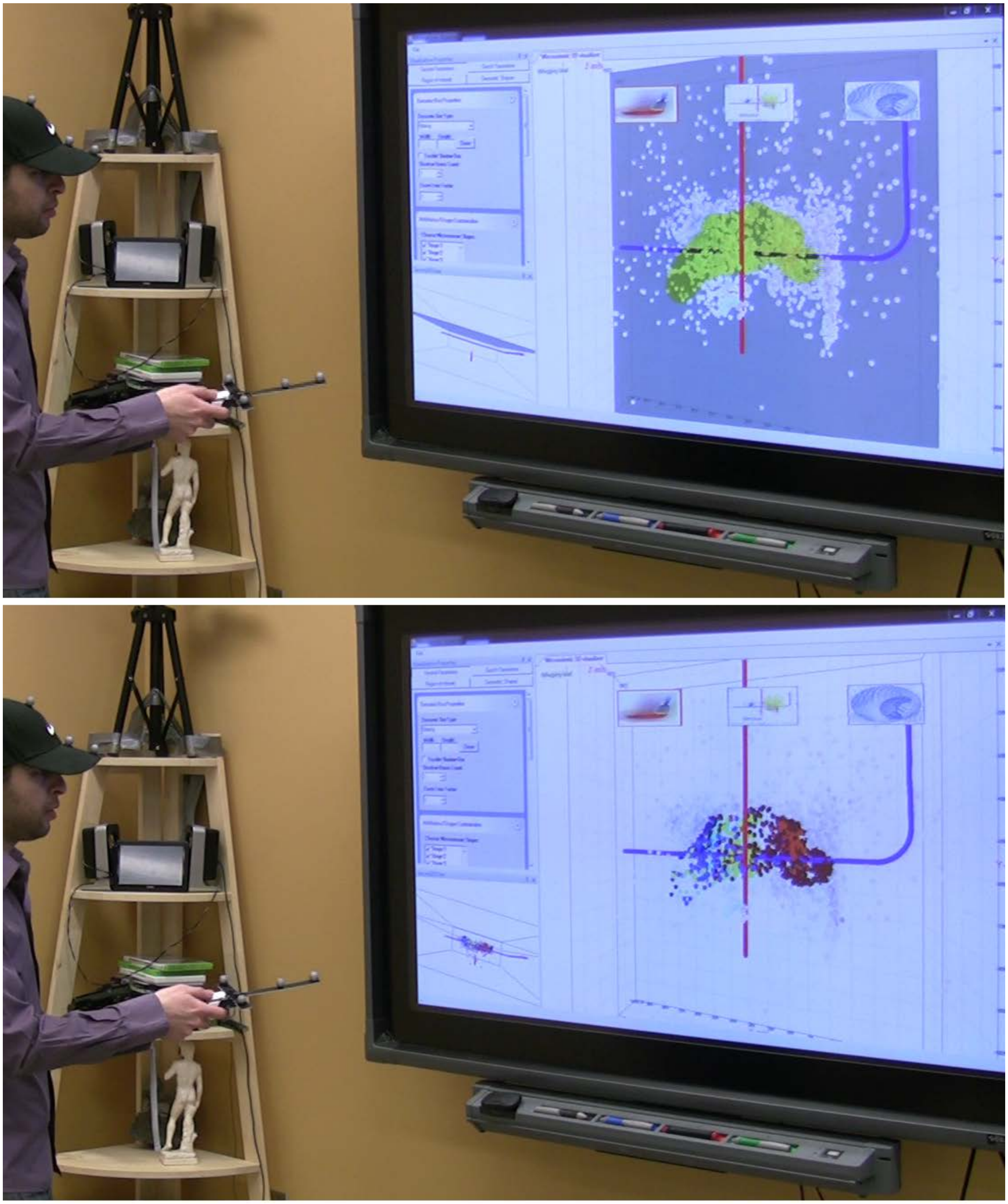

Figure 4.11: Brushing the painting surface (top), results of painting (bottom) .

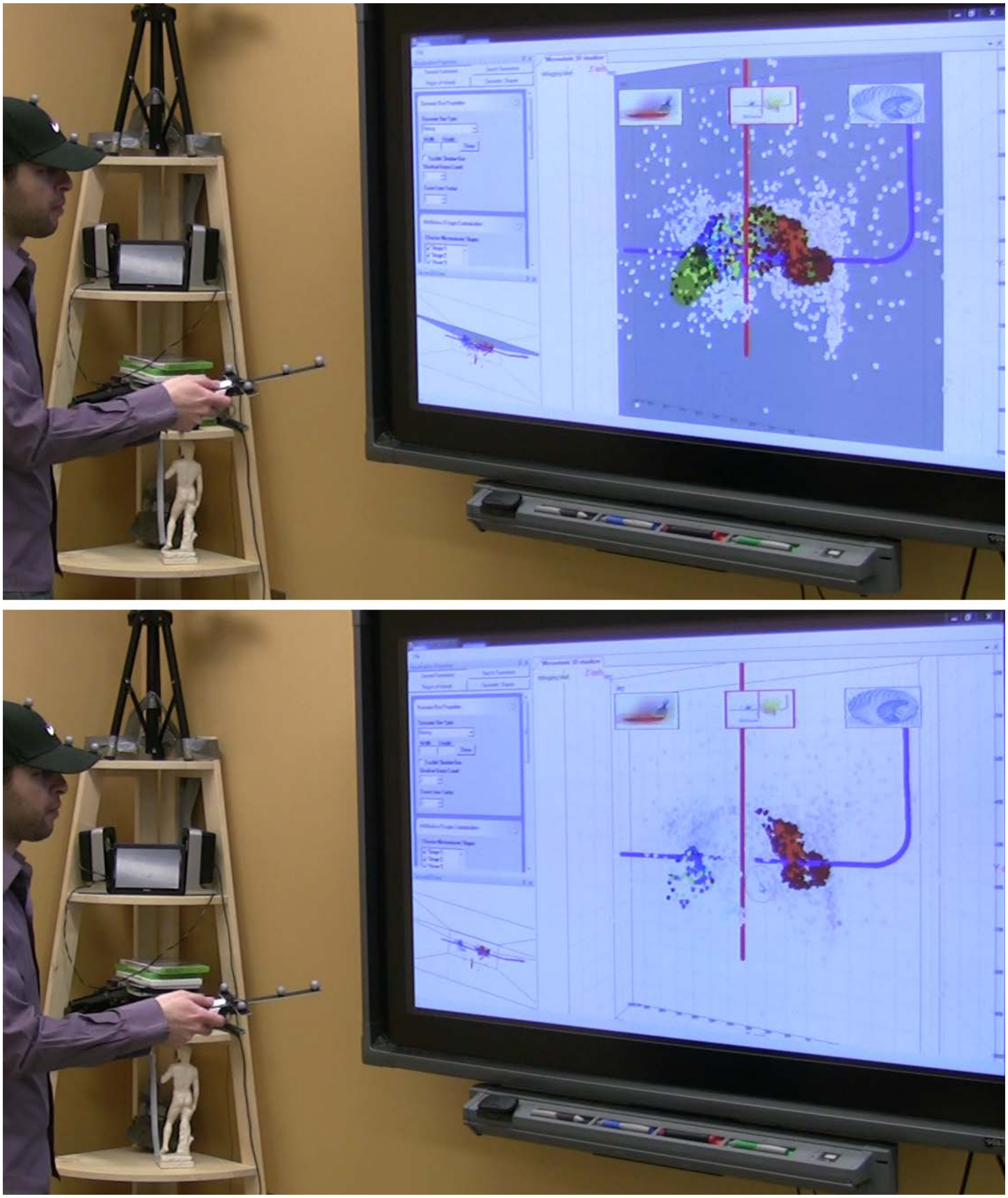

Figure 4.12: Modifying painted area to refine subset selection (top), results of painting after refinement (bottom).

field deployment, an issue that may affect some user decisions. We expect a more cost-effective, scalable approach for motion tracking on commodity hardware, such as Microsoft Kinect, and by leveraging other capabilities of the WiiMote, e.g. its pointing capabilities for selection. This remains to be implemented and tested.

4.7 Future Work

We are continuously collaborating with the domain experts to understand their needs and processes in order to provide them with intuitive interactive visualization. While considering this work as an ongoing project, there are many improvements to follow.

First, we plan to extend/adapt our implementation to use Kinect as a simple and relatively cheap alternative for sensing the proxemic data. Second, while we used the Wii for pointing only, we think that it has more powerful features which could be used to further improve the interaction. For instance, by adapting the Wii as our virtual spray can for painting, we can improve the interaction by vibrating the Wii to indicate that the virtual spray can is almost empty. Third, visualizing 3D information is tricky and we are planning to adapt some rendering techniques to make it more realistic. For example, rendering spray particles to simulate the spray (painting) effect will improve the 3D visualization and make it more convincing. Fourth, Tangible and physical-based interactions provide a room for a more realistic interaction. For instance, empowering users to use different types of physical spray cans to manipulate and highlight different features of the data, would provide a more natural interaction. Finally, the interactions discussed here can be used in many future scenarios. For example, each user may use his own set of spray cans to interact in a different way allowing for collaborative exploration scenarios.

4.8 Summary

In this chapter, we have described our initial exploration regarding characterizing the 3D problems in the microseismic domain. Our goal was to improve interactions by domain experts when navigating and interacting with 3D microseismic data by combining proxemics and a spatially-tracked handheld pointing device (the Wiimote). In particular, we designed three interaction techniques: mapping a user's location inside the 3D world directly (proximity-based interaction), tracking the device's location relative to that world for fine-tuning the user's location (device tracking), and a painting metaphor (using the WiiMote as a pointing device) to facilitate data selection. We believe that each of them has the potential to be integrated with other 3D visualization systems to improve their interaction. Natural interaction, especially with 3D data, is still an open problem. Indeed, such interaction is also new to the domain of oil and gas even when they have their own visualization rooms. In the following chapter we present an exploratory visualization system to help analysis and exploration of petrographic multidimensional data.

Chapter 5

PetroVis: Petrographic Characterization and Visual Exploration

After presenting our characterization and visual exploration for the microseismic domain (Chapters 3 and 4), we now present our research with the characterization and visual exploration of petrography as another high dimensional oil and gas dataset and case study. In this chapter, we aim to present our efforts of applying scientific/information visualization and human computer interaction techniques to the domain of petrographic analysis in order to support petrographic experts while working with their highly dimensional complex data.

5.1 Overview

In this chapter, our main contribution is in the characterization of the petrographic domain problems including data description, tasks abstraction, challenges identification, and discussion of experts' processes and needs. We hope to encourage future research efforts to build and improve analytic tools for visual exploration of petrographic data. Our secondary contribution, *PetroVis*, is an interactive visualization prototype which we developed to support the visual analysis of petrographic data. *PetroVis* incorporates the use of interactive multidimensional visualization techniques coupled with statistical methods, aiming at providing integrated tools for the visual exploration of the data. We developed *PetroVis* continuously guided by the feedback and suggestions from our domain experts and collaborators. Furthermore, we detail the insight we found by observing our collaborator used *PetroVis* in real scenarios to perform data validation and qualification. Finally, we conclude by discussing our reflections and lessons we have learned during the design and development of *PetroVis* including guidelines for future improvements.

5.2 Methodology

We followed a structured task-oriented approach during the exploration of the domain of petrography and during the development of our visualization prototype. In our exploration, we attempted to replace existing manual approach and tasks, and our prototyping effort involved having a domain expert as an integral part of the design team.

In the process of exploring the high-dimensional domain of petrography, we collaborated with a reservoir-engineering student who has some industrial experience and who was available on-site (in our research lab) during the whole process of the collaboration. Our collaboration was not structured but ad hoc and occasional as we met with the domain expert whenever needed. Roughly, we collaborated for the duration of ten months twice a week. Furthermore, some of the collaboration sessions with our expert lasted few minutes while others lasted around three hours. In our collaboration, we were also inspired by observational field techniques such as rapid ethnography [66] where we; (1) observed the expert while performing the data analysis (manually), (2) took notes regarding the expert's environment and how it was organized, and (3) interviewed the expert regarding different aspects of the domain tools and the challenges She was facing. Four of the collaboration sessions were video recorded, with two of them being regarding the feedback of the developed visualization, while the other two focused on observing the work of the expert. Other two experts provided feedback by email about our visualization as detailed in the evaluation Section 5.5.

The raw data collected from the collaboration enabled us to achieved better understanding of the process of petrographic analysis and the major tasks involved with it. Through such understanding we identified the typical challenges faced by the expert and the ways of approaching the data (Section 5.3). Our approach was task-grounded through the clearly-defined task of "validation of petrographic data clusters", and we strived to focus our design to match the aspects of that task as well as general interactions to facilitate the correlation of the data minerals.

We created a visual exploratory prototype involving extended existing visualization using a real petrographic dataset that we have received from our collaborator. Interestingly, our collaborator was using the same dataset daily while trying to analyze the petrographic data. This enabled us to receive faster and more insightful feedback from our collaborator while refining our design and visualization.

5.3 Petrography Characterization

The process of petrographic analysis involves exploration and integration of very large amounts of data presenting many challenges in which high dimensionality is one of the major issues. Such process offers (unique) potential opportunities where we can take advantage of visualization & interaction techniques to simplify petrographic data analysis. In this section, we characterize the domain of petrography by describing the typical petrographic data structure, the major challenges, and common tasks. We hope that our characterization would encourage future information visualization efforts to enable better visual exploration of the complex petrographic data. To gather the raw data, we followed an iterative process inspired by observational and characterization techniques such as contextual inquiry [77] and rapid ethnography [66], in order to learn and explore the petrographic domain. We had continuous consultation sessions with an in-site domain collaborator (petroleum engineering student), who had previous industry experience and who was part of our research lab for more than ten months. Our collaborator provided us with extensive support and insights regarding the petrographic domain.

5.3.1 Data Description

Petrographic data is usually represented as a high-dimensional database detailing rock samples from some geological basin (a basin is a depression in the crust of the Earth in which sediments accumulate). In our work, we experimented with four different petrographic datasets representing the description of rock samples from four different basins. The compositional structure of all the datasets includes 280 thin-sections (microscopic samples) collected from 30 wellbores. The

	Well 1_1	Well 1_2	Well 1_3	Well 2_1	Well 2_2	Well 2_3	Well 2_5	...
Depths_well	1747.6	1757.4	2081.6	1232.7	1233.7	1237.8	1264.6	...
Quartz.monocrystalline	28.3	37.2	18.8	40	26.6	33.2	35	...
Quartz.polycrystalline	3.1	6	3.6	3	1	1.4	2.6	...
K.feldspar	15.3	12.8	8.8	13.6	15.5	10.5	11.4	...
Plagioclase	2.3	0.8	0	1.5	1.1	1.8	1.2	...
Biotite	1.1	1.2	2.4	1.1	1.5	5.8	5.3	...
Garnet	0.4	0.4	1.2	1.5	5.7	2.2	3.8	...
Lithics.fragments	16.8	7.6	19.6	1.1	8.6	6.1	3	...
Grain.replacive.clays	0	0	0	0	0	1.5	0	...
Kaulinite	8.8	14.8	7.6	5.7	0	0	6	...
Chlorite	0	0	0	0	0	0	0	...
Calcite	0	0	0	2.8	27	31.8	9.8	...
Dolomite	7.3	5.2	0	0	0	0	0	...
Authigenic.quartz	1.5	0.8	2	1.1	0	0.7	0	...
Pyrite	0	0.8	0.4	3	8.6	2.2	1.5	...
Others	0	0	0	0	0	0	0	...
Intergranular.porosity	8.4	5.2	1.2	19.6	0.3	0	12	...
Intragranular.porosity	4	3.2	0	1	0	1	0.4	...
Mouldic.Porosity	2.3	2.4	0.4	3	0	1.8	6.7	...
Fracture.Porosity	0.4	1.6	0	0.5	0.4	0	0	...
Oversized	0	0	0	0	0.7	0	0	...
Shrinkage	0	0	2.8	0	0	0	0.4	...
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
CLUSTER	1	1	2	1	2	2	1	...

Figure 5.1: The typical structure of a petrographic dataset of some basin.

columns in each of the datasets correspond to the thin-sections from the cored wellbores, and the rows correspond to the attributes (e.g. minerals) observed through the microscopic analysis & measurements. Each sample in the database layout (Figure 5.1) has a name which is composed of the name of the well along with a number, representing the sample according to its sampling depth. For each well, the bigger this number, the deeper the sample is. We can think of the sampling depth as the key identifier of every data sample (the primary key of a data table if we are using database terminology).

The insight shared by our domain collaborator highlighted unique characteristics about the petrographic data. First, the expert stated that some petrogrpahic attributes, such as "permeability" and "porosity", are more important than others because they reflect correlation between reservoir characteristics, and their understanding facilitate building reservoir models. Secondly, some petrographic datasets are very detailed and some of the minerals may exist in different ways in the thin-sections leading to different levels of detail while generating the data. In particular, the level of detail of the data differs according to the method of the microscopic analysis and its accuracy. For instance, one representation of a mineral such as “Quartz”, within one dataset, is described as one record (line), but described as many records in another dataset (Figure 5.1 shows the attribute of “Quartz” as “Quartz.monocrystalline” and “Quartz.polycrystalline”). Interestingly, such characteristic may intensify the challenges of data interpretation since the dimensionality of the data increases proportionally to the number of minerals.

5.3.2 Task Analysis

The analysis of petrographic data can be carried out for the data samples of one wellbore or for the samples coming from multiple wellbores, thus two types of analysis exist, intra-well and inter-well analysis. Intra-well analysis focuses on the analysis of attributes and samples within the same well. Inter-well analysis, on the other hand, involves analyzing samples and attributes among multiple wells. From another perspective, the process of petrographic data analysis generally includes different phases involving many tasks. The first phase involves the generation of petrographic characteristics through the microscopic examination of the gathered samples. The second phase incorporates the use of statistical packages including clustering algorithms to generate the petrofacies (sample clusters). In the third phase, petrofacies analysis is performed including properties correlation as well as validation and qualification of the petrofacies. Such analysis is usually carried out manually due to the lack proper computational tools. Clearly, the expert’s knowledge is needed in order to try to make sense of the results before building the actual prediction model. Finally, the experts try to build a prediction model which would better characterize and explain the behavior of the reservoir. Next, we detail the (manual) tasks of clusters validation and qualification including the flow of their operations and how they are achieved by the experts.

Before the manual validation of the petrofacies (clusters), and in a typical clustering scenario, the expert would run a script in the statistical package over some input dataset in order to generate the petrofacies. After that, the expert will check the samples in each cluster and compare them and their properties to assure that they are not misclassified. Furthermore, such analysis would involve trying different clustering methods and/or tuning the parameters of each clustering method as well as correlating their results in order to optimize the generation of the potential petrofacies. In fact, clustering is one good example of an operation that could be simplified. Interactive visual exploration of the clusters can be very helpful by harnessing the cognitive power of the experts leading to faster analysis.

Qualification of the data is another important task for understanding the distribution of samples as well as the correlation between the data attributes and the petrofacies. Our domain collaborator described performing qualification by examining the petrofacies and the samples of each well, and then comparing manually the results from other wells in order to classify the petrofacies. In other words, qualification involves correlating the minerals in order to identify which of them affects certain petrofacies. For instance, qualification may enable grouping all the petrofacies who possess high amount of cement together. We observed that this process can be automated and interactive visualizations can be effective in highlighting the potential correlation and similarities in the data.

Building prediction models that would help in explaining the distribution of heterogeneities of the reservoir will lead to better characterization. However the task of building predictive models is complicated and requires creating mathematical models as well as adapting machine learning algorithms for classifying the data samples. Al-Anazi and Gate [8] explored the capability of support vector machines to classify lithology from well logs. Our data comes from thin-sections and does not come from well logs but the general process is very similar and can be extended easily. Indeed, in appendix A, we present a side-effort to explore the model of support vector machines to classify petrographic data. Interesting, our preliminary results showed the potential of exploring the data using computational techniques. Taking this further, we observed that integrating such model with petrographic visual analytic systems can simplify the task of qualification and ultimately support easier exploration of the data.

5.3.3 Understanding Domain Challenges

Working with petrological data and performing petrographic analysis is affected by the major challenge of the high dimensionality of the data. In fact, the minerals identified from some basin may have 400+ dimensions. Furthermore, the challenges associated with the analysis of this complex data are being intensified due to the use of manual methods and the lack of proper geological tools until very recently [24]. The insights shared by our collaborator highlights that analyzing petrographic data is time consuming and requires dedicated experts.

Although experts started to use computational tools (including statistical methods and machine learning algorithms) to help them analyze this vast amount of the data, there is still lack of effective tools to simplify the analysis process and make it less manual. One example of this lack of computational tools can be explained through the following scenario: one domain expert may analyze the data by applying some clustering methods to learn more about the similarity and grouping of the data samples. However, different clustering methods yield different clusters and the experts can not easily compare and validate the resulting clusters. Another example of the limitations faced by experts while analyzing the data is the manual data qualification. Qualification of the data involves manually correlating different samples and minerals among multiple wells as well as manual identification of the minerals that affect certain clusters. According to our expert, this process is time consuming and comparison of many samples, petrofacies, and minerals can be error-prone since most the operations are handled manually. We believe that interactive visualization can be of great help to support petrographic experts and aid the validation and qualification of their results.

Interestingly, experts face another type of difficulties because of certain data characteristics. One example of such difficulties exists due to the different levels of detail of the data. As we described in Section 5.3.1, experts attribute that difference to the method of the microscopic analysis leading to high heterogeneity of the data. Additionally, some attributes may affect indirectly other attributes, thus increasing the data dependency. For example, the porosity of the rock depends sometimes on the values of other minerals which may grow/shrink overtime affecting rock void spaces and thus the porosity. Such hidden dependency makes it difficult to correlate the minerals. Therefore, experts are demanding visual analytic tools to help them explore their complex data and discover such correlations.

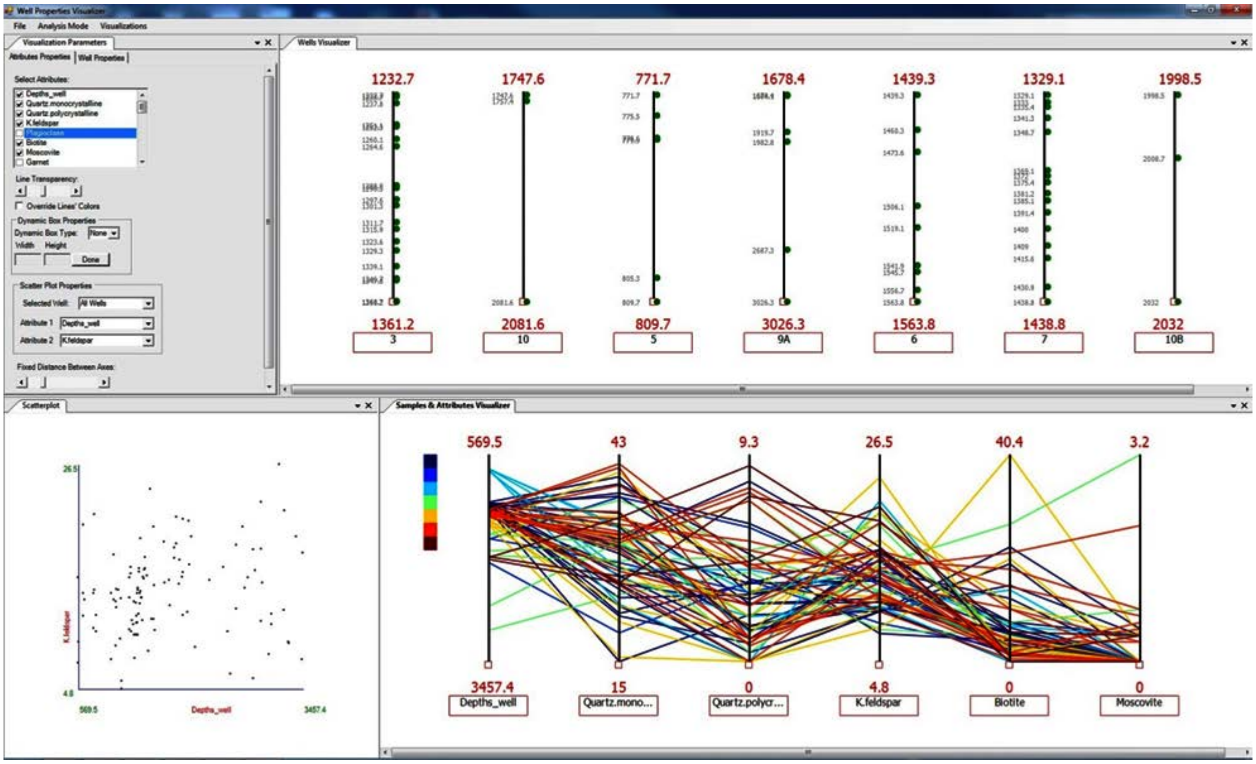

Figure 5.2: Overview of the *PetroVis* visualization showing wells and samples coordinated visualizations.

5.4 PetroVis

We designed and developed *PetroVis*, an interactive exploratory visualization prototype to support petrographic data analysis (Figure 5.2). *PetroVis* consists of a set of visualizations organized through two analysis modes. The first aims at helping the experts analyze visually the data samples and correlate the attributes. The second provides advanced data correlation by coupling statistical methods to extend the parallel coordinates visualization in order to support intra-well analysis. In this part, we explain our design decisions, visualization techniques and visual encoding. Then we explain the different components of the system with emphasis on the developed interactions.

5.4.1 Design Decisions

PetroVis is designed keeping in mind the earlier developed characterizations (Section 5.3). We carefully chose visual variables [21] (such as color and shape) which effectively represent and highlight the important features of the petrographic data. For instance, we decided to represent each well using one color and we used the same color for all the samples belonging to that well. Clearly, it became easier to relate which samples belong to which well just by observing the color of the sample. Similarly we decided to encode petrographic clusters according to a color map in order to correlate each cluster and its associated samples. In addition, we used a circle (along with an attached number) to represent a sample and its associated depth. The position of the circle in our visualization reflects the actual position of the sample depth.

Our choice to visualize the petrographic data was highly influenced by the structure and the dimensionality of the data. Besides, domain analysts are usually familiar with scientific tools such as Matlab and they are familiar with cross-plotting techniques for correlating two or three data attributes. Accordingly, we decided to use and extend the technique of parallel coordinates (PC) [46] to support the visual analysis of the high dimensional petrographic data. The typical visualization of PC represents each data attribute (in our case mineral/feature) as parallel spaced lines (axes/columns) where each line is represented by its name, the min, and the max values of all the samples. Each data element (in our case well sample or thin-section) is being represented as a polyline intersecting every (visible) data axis at a position proportional to its value for that dimension. Although PC suffers from data cluttering, we tried to simplify this issue by adapting certain strategies such as data filtering and axis reordering [76]. It is worth noting that we did not experiment with other multi-dimensional visualization technique because we wanted to focus on the use of PC, but we are not sure if other techniques may be better fit or the visualization of high-dimensional petrographic data or not.

In our approach, we decided to support different data filtering interactions to enable the experts to filter out certain samples and analyze only those which are coming from specific wellbores or that satisfy certain depth range. We also designed a statistical-based axis reordering interaction in order to evaluate the importance of each data attribute according to some statistical method, and reorder them by showing the most important ones first. Similarly, we introduced a master/detail visualization to facilitate attributes correlation especially those who are composed of sub-attributes (different levels of detail). In addition, in our visualization we decided to support other interactive features such as details-on-demand [85], color maps and scatter plot [28] to facilitate data exploration. Finally, we adapted interactively the technique of multiple coordinated views in order to synchronize the different visuals and simplify the data analysis. According to our domain collaborators, our design choices make sense and effectively represent the data aspects.

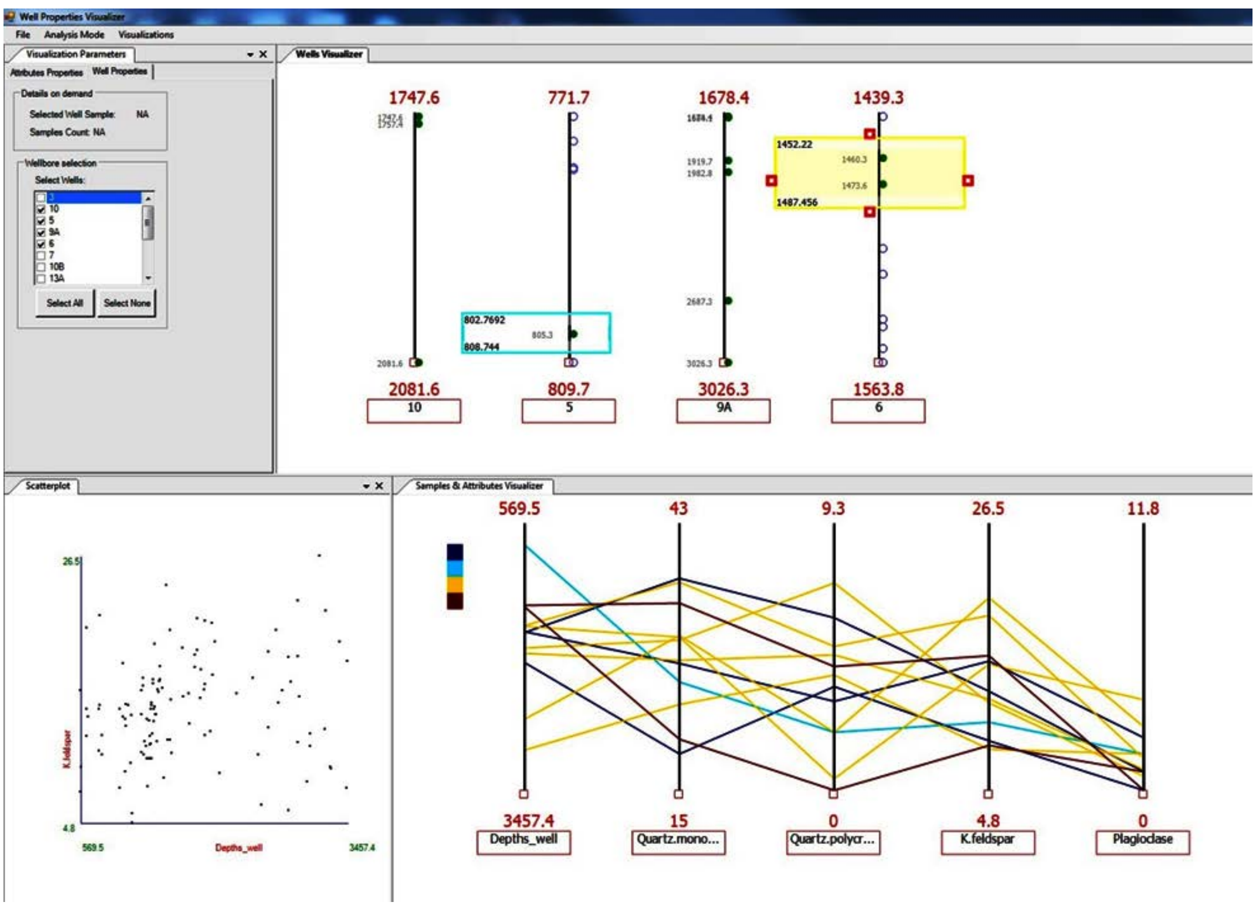

Figure 5.3: Filtering data samples from two wells to refine the subset for analysis.

5.4.2 Visualization Components

The main interface of *PetroVis* shows an overview about the petrographic data through the coordinated visuals seen in wells' visualizer and samples' visualizer (Figure 5.2 top and bottom respectively). The user can customize which wellbore to consider while analyzing the data and also which minerals (or data attributes) to focus on. Wellbores are being visualized as vertical parallel columns where each column represents one well. Each well is also represented as one color slot in a color map shown in Figure 5.3 bottom. Well samples associated with each well are being visualized by green circles rendered according to their depth values (Figure 5.3 top). In addition, the user can interactively filter the sample set by selecting a subset of the samples associated with one well or all wells (Figure 5.3). This coordination is happening automatically so that when the user removes some well(s) and/or sample(s) from the wells visualizer, the PC visualization will be updated automatically to reflect that.

Within the PC visualization, the expert can perform axis reordering manually to correlate the minerals, and/or brush (filter) the data samples in order to refine the samples subset. In addition, the user can inspect the details of any well sample by hovering the mouse cursor over its polyline, then the visualization will show the values of the nearest minerals for that sample. This feature enables quantitative analysis of the data and eliminates the need for the analyst to go back to the database file to check it. Figure 5.4 shows how the user selected only samples from the third wellbore (Well-3) (by clicking over the dark-blue color entry in the color map), and how he used the mouse pointer to inspect the values of the minerals "Moscovite" and "Biotite" for a certain sample. Finally, *PetroVis* supports the visualization of scatter plot along with the existing visuals to give the experts the opportunity to map familiar visuals with new ones, in a hope to simplify learning the new visualization.

PetroVis also provides intra-well visual analysis in order to support petrofacies validation and qualification. Within the mode of intra-well visualization, the analysis is carried out for each well individually through two coordinated (enhanced) PCs 5.5. The two PCs are visually related; the

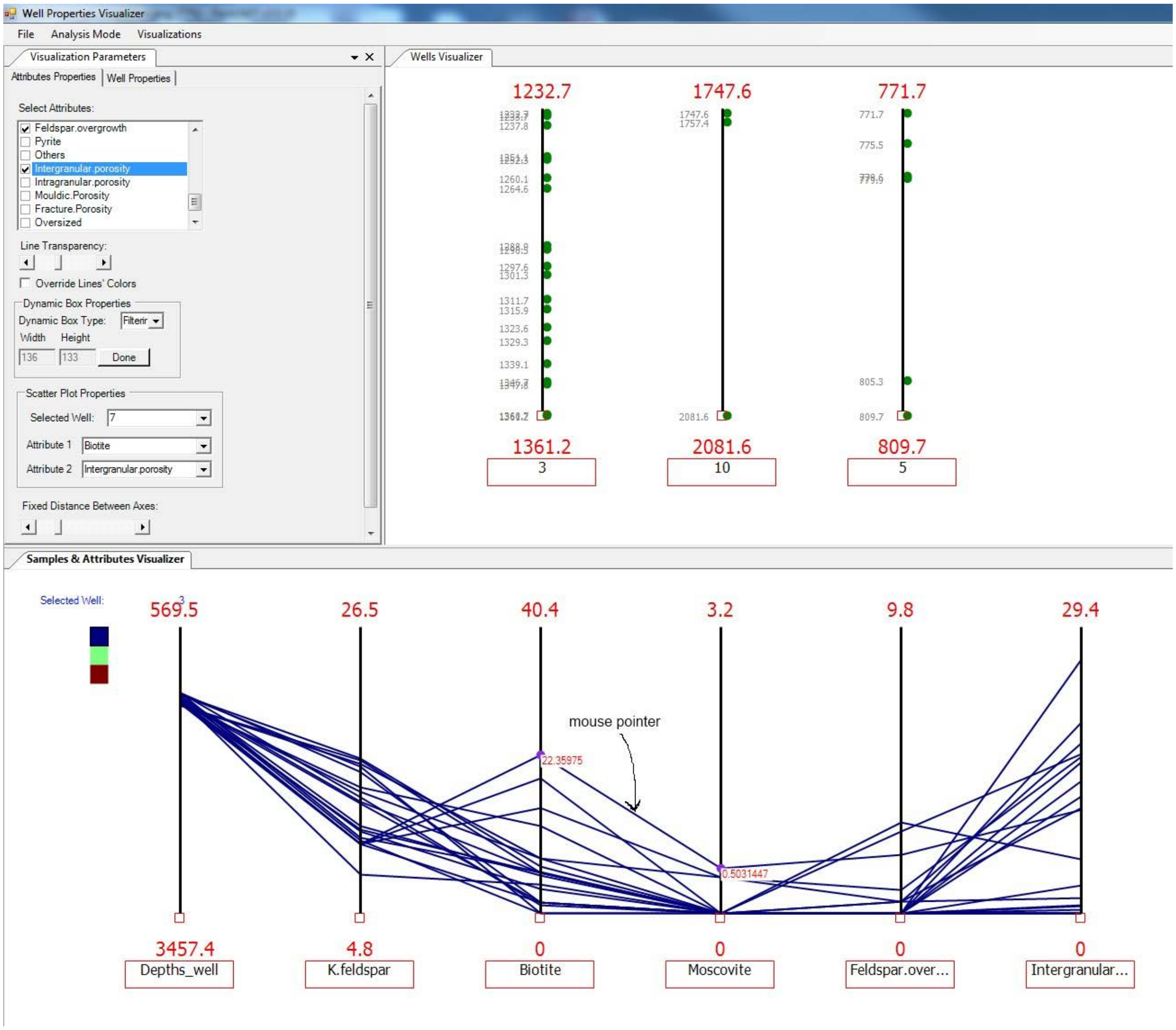

Figure 5.4: Inspecting sample values while focusing the visualization to only show the samples from well-3.

first one will show the original attributes (expanded attributes as in Figure 5.6 bottom) and the second one will show the attributes in a combined way (collapsed attributes as in Figure 5.6 top). Any combined attribute is defined as follows: the sum of the percentage of each (expanded) attribute representing the same mineral. Figure 5.6 provides an example for the mineral “Oversized.Pore”. In the original description, Oversized.Pore occurs in three different ways: “Oversized.Pore as monomineralic grain”, “Oversized.Pore as monomineralic grain deformed” and “Oversized.Pore in plutonic rock fragment” (the expanded attributes). The combination of these attributes results in

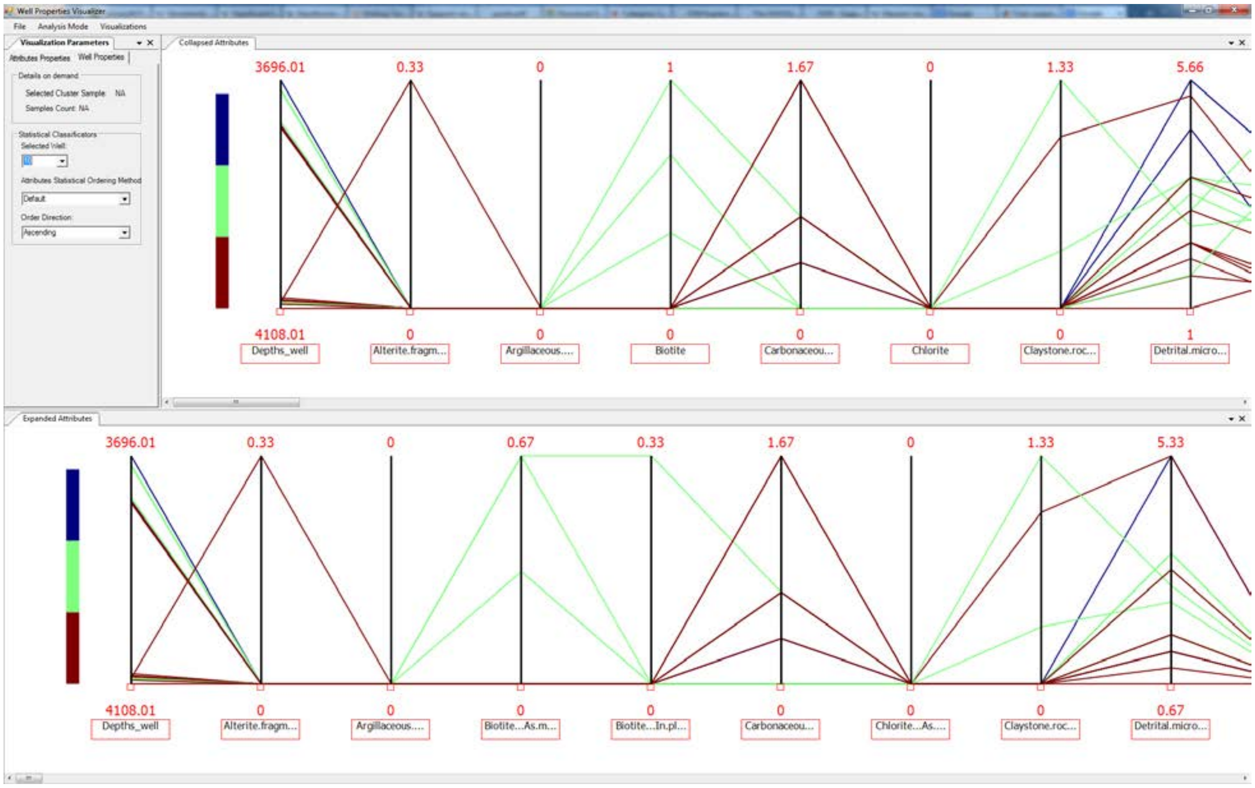

Figure 5.5: Overview of Intra-well visualization showing collapsed and expanded attributes.

the more general attribute Oversized.Pore (the collapsed attribute). In our visualization, the user can select one of the combined attributes and the other coordinated visualization will automatically highlight the expanded attributes, if any, for that selected attribute. In addition, colors represent data clusters. Hence samples with the same color theoretically represent the same petroface for that well as indication of sharing similar characteristics.

We also extended the axis reordering of the PC through statistical methods including Variance, Inter-Quartile Range, and Range (Max-Min) in order to override the default ordering of the data attributes and organize them according to their importance (Figure 5.7). Furthermore, the user can select whether the data should be sorted in ascending or descending order, and the user can perform this interactive reordering per well or per cluster if needed. Our visualization prototype fully supports intra-well analysis, but we believe that there is a real potential to extend this work further in order to support inter-well analysis, to correlate and qualify the data among multiple wells. Indeed, we are considering this as part of our future work.

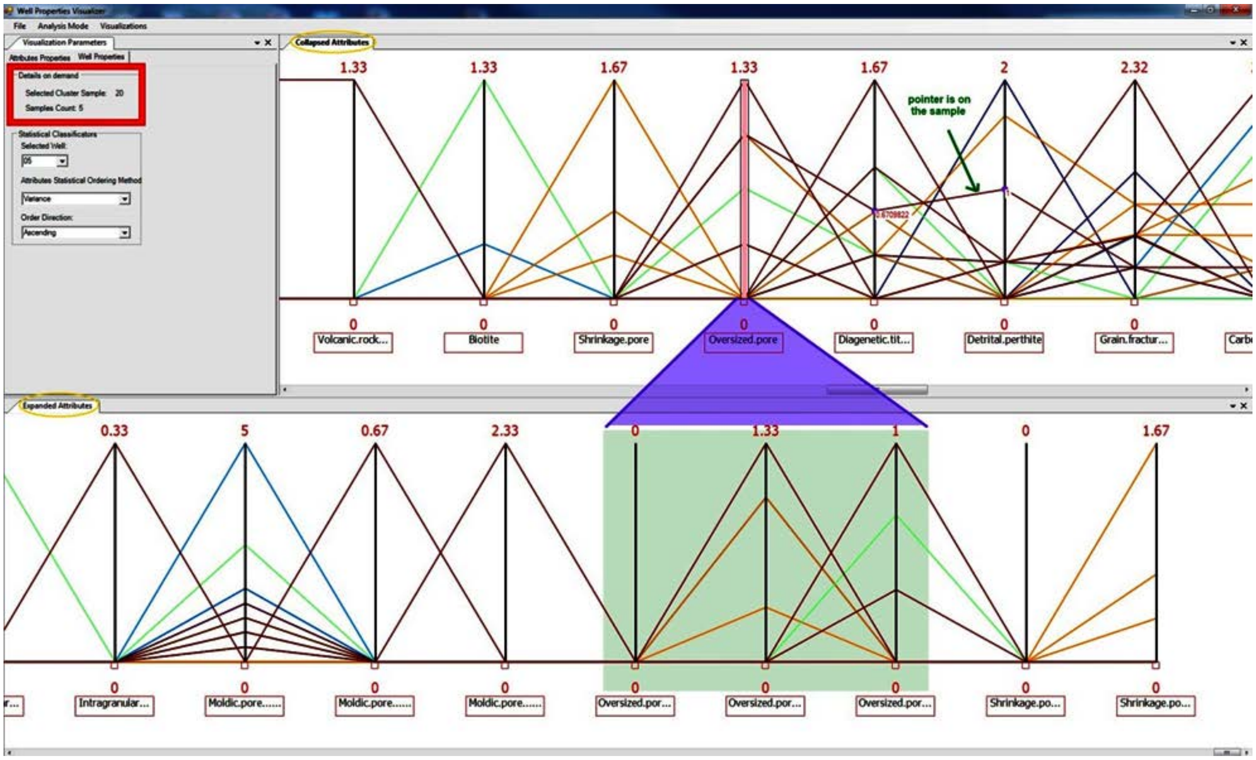

Figure 5.6: Mater/Detail visualization showing how collapsed and expanded attributes can be linked together .

5.5 Evaluation

We designed and developed *PetroVis* continuously with our domain collaborator. We also conducted post development evaluation sessions in order to validate and reflect back about the usability features and limitations of our tool. We had three domain experts as our participants (P1, P2, and P3). P1 is our on-site domain collaborator who provided us with continuous insights and suggestions during the design and development of *PetroVis*. P2 is an expert geologist with high experience of working with petrographic analysis, and who performed the manual analysis for some of the datasets which we used. Finally, P3 is a reservoir engineer with extensive industry experience who evaluated our tool and provided detailed feedback. Apart from the on-site domain collaborator (P1), the other two specialists provided their feedback through email. We sent them our tool with a tutorial detailing how to use and interact with it. After that, they tried the tool for one to two weeks and then provided their feedback to us. In this section, we will present

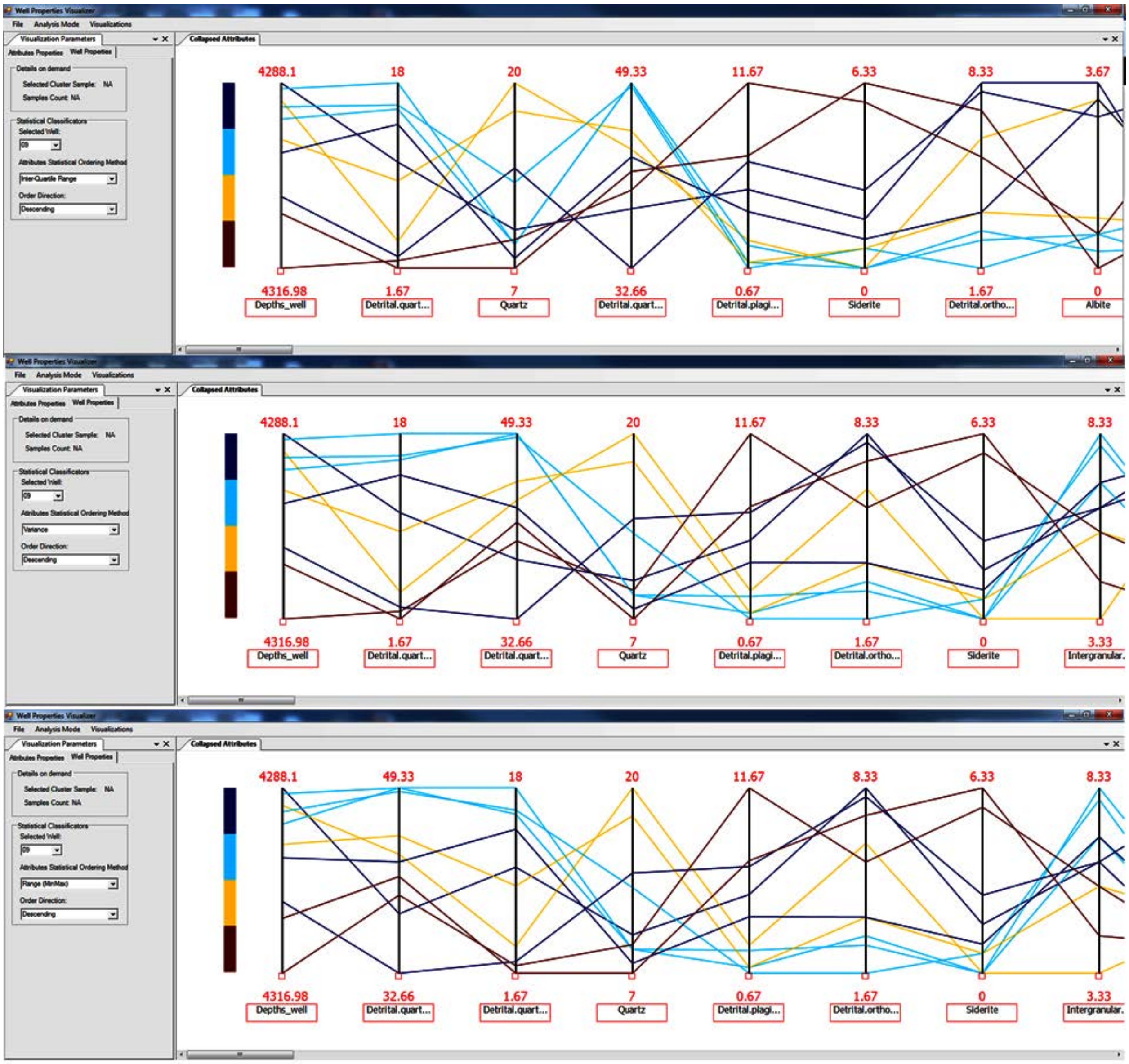

Figure 5.7: The result of applying different statistical-based reordering methods to refine the ordering of the attributes within the PC visualization.

three threads of evaluation that we have performed; usability evaluation, case study evaluation, and illustrative example evaluation.

5.5.1 Usability and Interaction

During and after the development of our prototype, we received feedback about the usability features of *PetroVis* as well as limitations and suggestions for improvements. Actually, we still believe that further and deep evaluation of *PetroVis* should be conducted in order to confirm its effectiveness for analyzing petrographic data. In this part, we present the results of the informal qualitative analysis that we performed.

What Went Well?

Our participants who worked with *PetroVis* provided positive feedback about how our tool possesses the potential to support petrographic data analysis. In particular, P2 expressed *"It is possible to interpret what the diagenetic process made in the depositional history of the well because you can correlate any diagenetic attribute with the depth (which can tell about the changes in rock structure and/or behavior over natural factors such as time, temperature, and pressure)"*. The expert added, *"A big importance is that we are able to (exactly) visualize which minerals control the petrofacies"*. This last feedback shows how *PetroVis* enables insight discovery regarding minerals and clusters correlation. Another expert (P1) provided feedback regarding the feature of attributes reordering. P1 expressed *"reordering the attributes based on their importance is really important since it minimizes the error of the analysis when guided by the geologist's knowledge"*. Additionally, P1 performed intra-well analysis using *PetroVis* to validate and qualify the petrofacies (See Section 5.5.2 for more details), and he/she observed that the feature of statistical-based reordering is useful for doing this type of analysis. In fact, after P1 performed clusters validation, P1 was able to correctly confirm most of the samples in each cluster and has found few samples that have been misclassified (More on this in section 5.5.2).

Limitations

Regarding some of the limitations reported by our participants. P2 expressed a first impression about our tool saying it looks quite complicated and a tutorial is needed to explain the interface. P2 specifically added that *"what is causing more doubts is the fact that it is missing several interface captions. It should have something 'explaining' the interface elements such as a label for each color map entry*". In addition, P2 highlighted that it is not clear which dataset is being analyzed while working with the tool, since there is not any place telling this information in the graphical interface. P1 and P3 reported certain bugs with our current prototype, and they detailed the steps to reproduce these errors. Accordingly, we fixed some of the mentioned errors, and we decided to re-design some of the features which would automatically fix the others. P3 also reported that the tool does not provide a way to export the results of the analysis such as capturing and saving the graphs which is sometimes needed to continue the analysis later.

Ideas for Improvements

Other participants reported different suggestions for improving *PetroVis* as well as minor issues they observed, and it would be better if these issues have been handled. P2 suggested that the "sample-depth" values to be rearranged such that the minimum value is shown at top of each visual axis to match the expected physical sample depths of the well. Another expert, P3, highlighted the importance of analyzing extreme values. Specifically, P3 suggested to put an option to enable manual tuning of the 'Min' and 'Max' values of each attribute. The expert added, *the reason is that the min and max of a property is not necessarily reflected in the data, and this might provide unclear indication of the property cross-correlation. In essence, the geologist or the geo-engineer should be able to find the cluster extreme values*. In addition, all of our participants are familiar with statistical measures, and they expressed the importance of seeing these features within the visualization. For instance, they expressed that it would be extremely helpful for geo-engineers to embed some statistical measures for each property such as mean and variance. Finally, most of the experts also pointed out the issue of inaccurate sample selection using the mouse, especially when there are many cluttered samples. Therefore, they suggested adding the option to manually interact and define them.

In *PetroVis*, one participant (P3) commented about our scatter plot visualization saying:

it is highly beneficial to color the data points according to a third variable. Doing so may enable some sort of clustering to be visually captured.

P3 also expressed that it would be a very good attempt on the cross plot to allow the user to interact by the mouse and select the data attributes and filter the data, and the PC should hide or shows just those ranges as an automatic filter. The same expert continued, this is highly important for seismic or well-log analysis where multiple data are available and each cross plot could give the experts an indication of the (petro)facies. Another participant (P2) also highlighted the importance of scatter plot and having it coordinated with the other interactions. For example, if the user selects a well (from the options menu), then the range of the values within the scatter plot should be updated to reflect that selection.

In essence, all participants liked *PetroVis* and expressed that it can be useful for analyzing the petrographic data. They also highlighted some limitations and suggested many ideas to further improve the current visualization and make it more effective for petrographic visual analysis.

5.5.2 Cluster Validation

In this part we explain how our tool has been used as part of a real task (case study) to validate and interpret the petrofacies results. The task of cluster validation aims at confirming the validity of the resulted clusters, the distribution of the samples, and the effectiveness of the clustering method (partitioning or hierarchical). Following we describe a walk-through detailing how our domain collaborator has used *PetroVis* to validate the data clusters.

P1 used the mode of intra-well analysis of our visualization to perform cluster validation. At the beginning, P1 started analyzing Well-1 (statistically) by applying PAM and AGNES as (partitioning and hierarchical) clustering methods to identify the petrofacies of that well. The results show three found clusters. However, P1 highlighted that the resulting clusters, from each method,

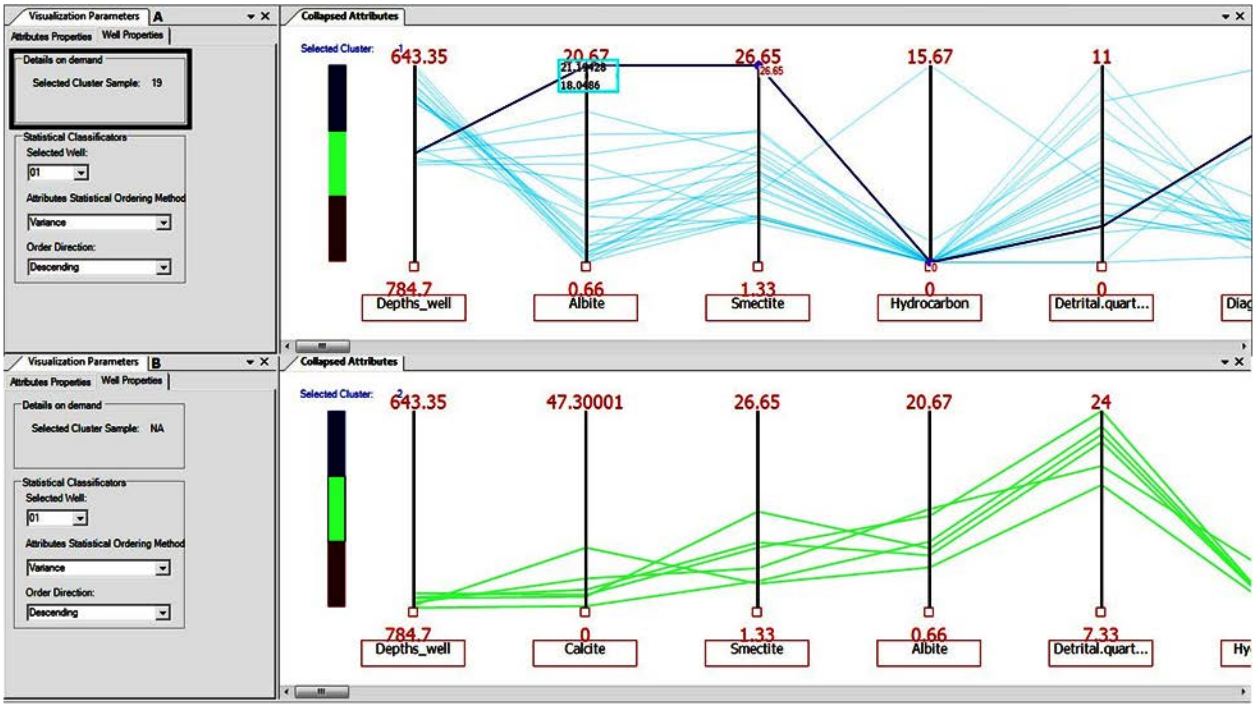

Figure 5.8: Identifying which minerals would explain the behavior of having the sample "Well-1-19" clustered differently (top), and examining the well-defined trend of the second cluster (bottom) .

are not similar and there is a need to validate them and hopefully understand why some differences exist. Initially, P1 identified the samples "Well-1-19" and "Well-1-27" as being possibly misclassified. Following, P1 used *PetroVis* to visually examine the clusters and their associated samples, and to try to find out which minerals affect each cluster and which clustering method is more accurate.

According to P1, the details of the visual analysis are as follows. Applying the Variance method (over cluster 1, the dark blue in Figure 5.8-A) caused the minerals to be reordered according to their importance. The expert added, the sample "Well-1-19" exists in one cluster of the PAM result, while it appears in a different branch (of a different cluster) in the AGNES result; and the PC visualization clearly exhibits these characteristics. In fact, the PC shows that the minerals "Albite" and "Smectite" are the possible reasons for such behavior (Figure 5.8-A). On the other hand, Figure 5.8-B shows that the cluster 2 (the green cluster) is well defined and exhibits a well formed trend. Regarding the 3rd cluster (the brown in Figure 5.9), the expert tried to understand the behavior of the sample "Well-1-27". The expert expressed that this sample appears in a single cluster/branch in the PAM/AGNES results, and it was possible using the visualization to identify which minerals caused this 'different' classification (single-element cluster). P1 continued, filtering the PC visualization shows the sample "Well-1-27" containing high values for some of the attributes (Figure 5.9). The expert concluded, this may emphasize why this sample has been clustered separated in the PAM and AGNES results. In essence, such analysis reflects the easily accessible insight through the use of our visualization.

PetroVis has been successfully used by our collaborator for analyzing petrographic data and assisting the expert in the interpretation and validation of the petrofacies. Detailed analysis regarding cluster validation of the petrographic data using our visualization has been documented in the work of Cevolani et al. [24] where they developed a computational methodology to study heterogeneities in petroleum reservoirs.

5.5.3 Illustrative Example of Qualification

Qualification of petrographic data is an important task that requires the knowledge of geology and the meaning of the minerals. The goal of qualification is to relate and understand why certain minerals come together and how the existence of them affects others and ultimately affects the petrofacies. In general, the task of qualification can be performed for one well or between all the wells of one basin. In other words, this task could merge between intra-well and inter-well analysis. In this section, we will detail an illustrative example about how our domain collaborator used our visualization to qualify the petrographic data by assuming some knowledge about the aforementioned task.

Intra-well qualification aims to find the main attributes that differentiate one cluster from the others within the same well. According to our collaborator, this process involves a comparison between the clusters and is easily performed through the use of visualization coupled with statistical calculations over all samples of a well and over each cluster. P1 attempted to qualify Well-1, he or she expressed that the green cluster, which is a well-defined cluster as analyzed in the validation task (section 5.5.2), is mainly qualified according to the high values of the "detrital quartz

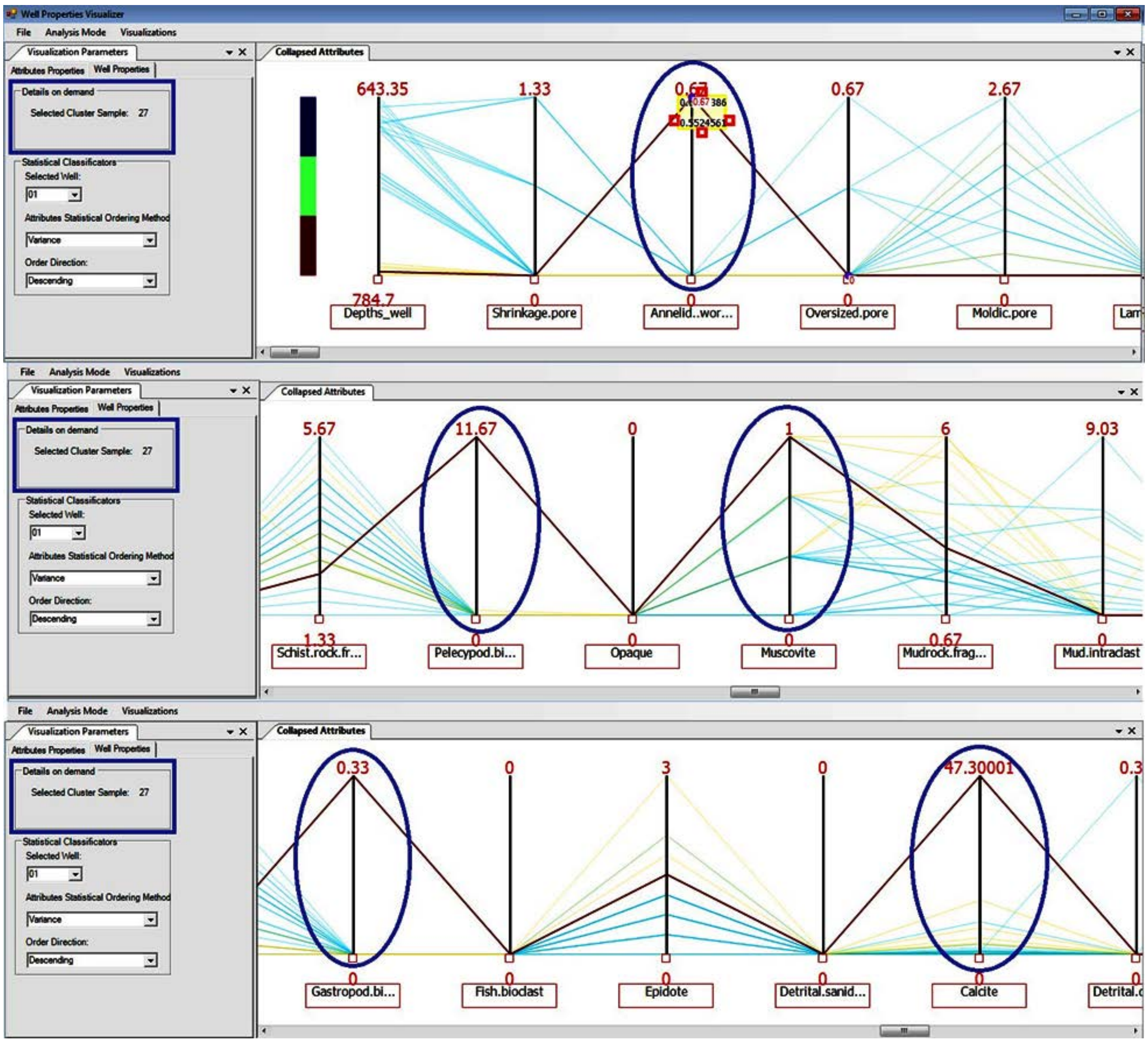

Figure 5.9: Identifying the misclassification of the sample "Well-1-27" by filtering it and highlighting the (high values) minerals that may affect its clustering.

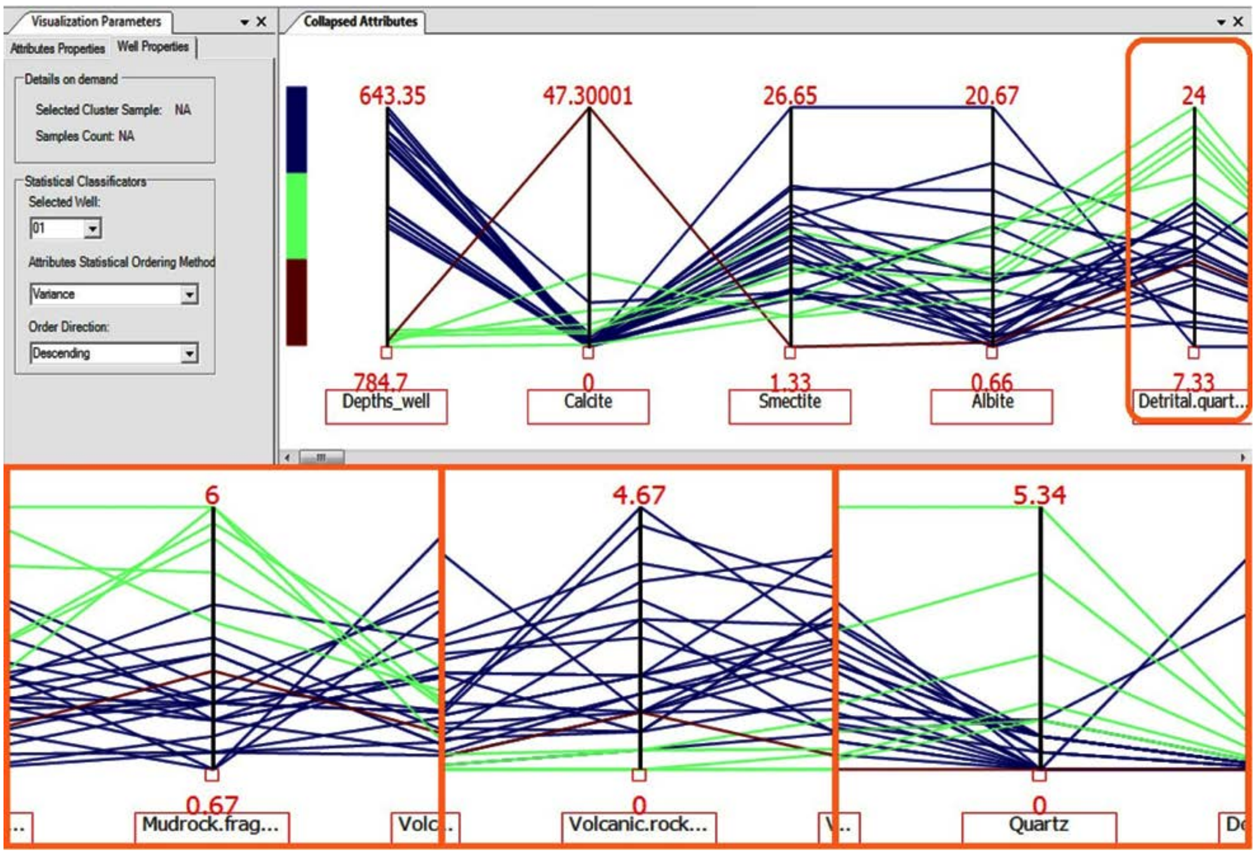

Figure 5.10: Qualifying Well-1 and highlighting the attributes that possibly affect the distribution of its samples.

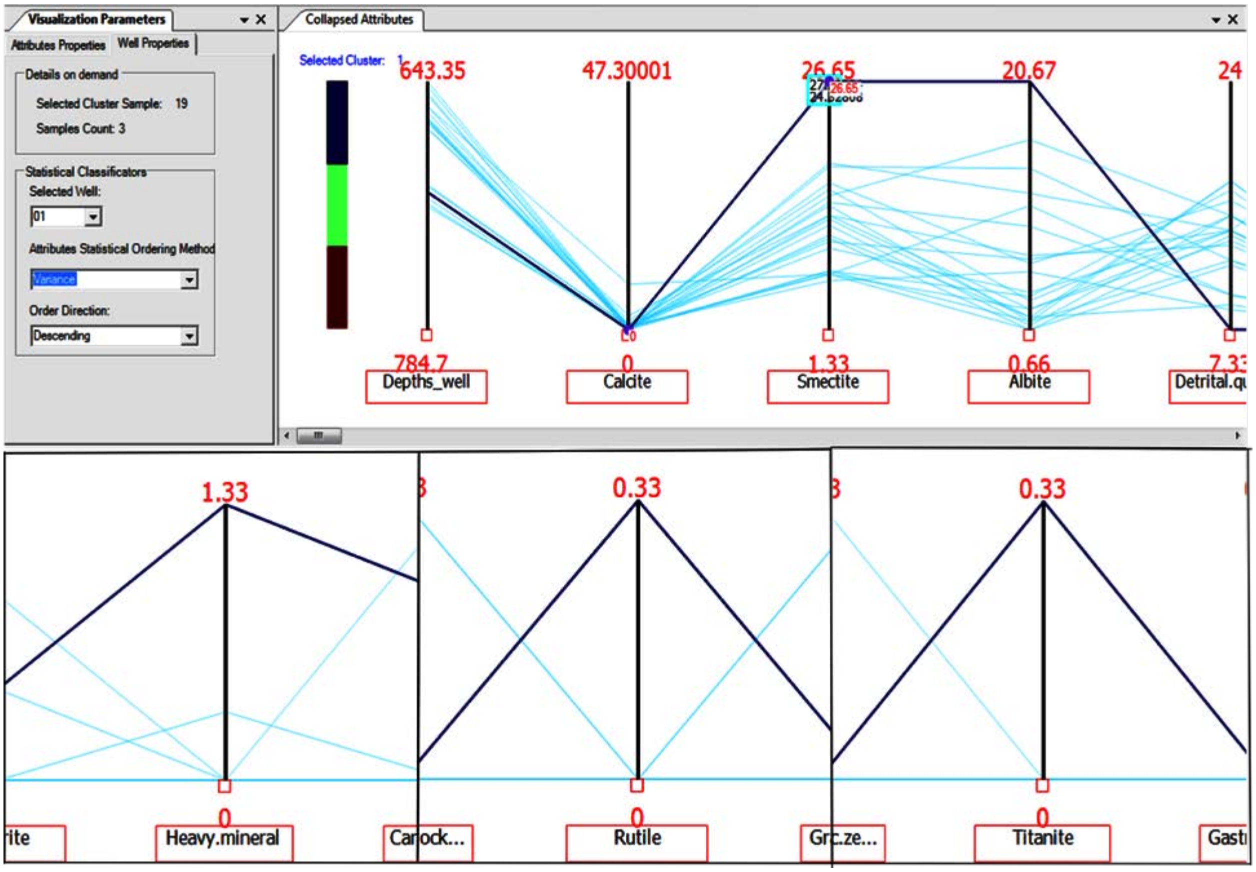

Figure 5.11: Using the PC to visualize which minerals affect the sample "Well-1-19" to understand its behavior.

monocrystalline" attribute (Figure 5.10 top). The expert added, other minerals that seem to have some importance in this cluster are "mudrock fragments", "volcanic rock fragment", and "quartz" (Figure 5.10 bottom). The other clusters (cluster 2 and 3) have been qualified similarly. P1 highlighted that qualifying the clusters may give intuition regarding the behavior of certain samples. For instance, the sample "Well-1-19", that presented problems in the validation task, has the presence of some attributes, such as "Heavy Mineral", "Rutile", and "Titanite" (Figure 5.11), but the remaining samples do not have.

On the other hand, Inter-well qualification aims to classify certain patterns in the data together by correlating (manually) the minerals from multiple wells. Interestingly, P1 commented that doing inter-well qualification was the most time-consuming task. P1 also mentioned that our visualization has been used only for some parts of this task, particularly for inspecting the minerals range for some of the samples in a try to compare and classify similar clusters together. Additionally, P1 specifically expressed that the most difficult part of this task was to compare clusters from different wells, since there was no way to see them at the same time.

Our tool currently does not completely support inter-well qualification, and we are considering this as part of our future work. In fact, P1 highlighted that doing inter-well qualification automatically is available through some statistical packages but without much interactivity. In addition, the expert compared the manual analysis results with the initial results of the automatic qualification from the statistical tool, and the results showed less similarity. P1 attributed this insight to the fact that while doing the manual qualification, he or she used the combined attributes. However, the automatic method applied the analysis for every single attribute with the others among all the samples. As a result, P1 suggested that it would be of great importance if the visualization can support showing multiple wells/clusters at the same time which would simplify the analysis and save the time.

5.6 Discussion and Lessons Learned

We had only a single domain collaborator available on-site during the whole process of exploring the domain of petrographic analysis. From our experience, we observed that building visual analytic system iteratively with domain collaborators is the direct way to insure its success. However, we stress that a clear understanding of experts' needs is crucial before starting any development. As visualization researchers, it was not easy for us to understand the petrography domain terminologies so we had multiple meetings with our collaborator where he or she explained the different domain aspects. Although having one expert only as our collaborator may be considered a weakness, it is not easy to find many expert collaborators, and it is still better than having many experts who are busy and provide less feedback most of the time. Besides, having that expert available on-site enables continuous real-time feedback whenever needed which was very valuable especially at the prototyping phase of *PetroVis*.

One of the experts highlighted that our visualization can be useful even for visually analyzing similar oil and gas data. He or she expressed that:

"Well logging would be a good candidate for your work. For example, we have a well and we measure different properties along the well (GR, RO, SP, Resistivity, ...). By selecting the high resistivity parts some intervals along the well will show up that might help define the nature of fluid in reservoir section using the other attributes (SP, ...)".

Such feedback reflects that our visualization can be adapted to other similar oil and gas instances. This feedback also reflects about the success of our visualization in conveying important ideas and helping the experts explore their high dimensional data.

Geo-engineering experts including our domain collaborator use statistical tools heavily in their work. Therefore, they are familiar with certain visualizations. In order to reduce the learning curve for such experts regarding our new visualization, we adapted well known existing visualizations (such as scatter plot) and we integrated them with *PetroVis*. Moreover, we enabled synchronization among all the visuals to make it easy to correlate any found insight from any visualization with the others. Finally, we customized the new visualization with many traditional interactions (e.g. data filtering) in order to facilitate the exploration of the data.

Finally, the results that we have received from the experts show that our visualization could be extended to potentially support other petrographic tasks such as inter-well qualification. Clearly, our collaborator highlighted the partial use of *PetroVis* while performing the qualification, and how it would be great if *PetroVis* is extended to fully support such tasks. Accordingly, we are considering this as a future work.

5.7 Future Work

Our visualization represents only a work-in-progress prototype and there many things to improve. Indeed, we received suggestions for improvements from the experts during the post-evaluation sessions that we have conducted. For instance, we plan to extend the scatter plots by integrating statistical features within the visualization and improve the synchronization with the other visuals. We also plan to automate some operations by integrating our visualization with some statistical packages to enable manipulating the data clusters on the fly. Finally, we plan to conduct a detailed formal evaluation of our visualization with many more domain experts in order to confirm that our system satisfies their expectations and simplify their data analysis.

Our collaborator also mentioned very interesting ideas to extend our visualization. Accordingly, we are planning to match the color maps that we are using in our visualization with the color maps used within the experts' statistical packages to make it easier to compare and correlate the results. We are also considering other intuitive interactions to manipulate and explore the data as part of our future work. The ability to visually manipulate the samples within a cluster and manually refine it is one example of improvement threads.

5.8 Summary

In this chapter, we explore the domain of petrography as a high dimensional space within the oil and gas domain. Petrographic analysis is important for characterizing oil reservoirs and increasing hydrocarbons production. We work closely with domain collaborators (e.g. geologists and geo-engineers) with the goal of helping them explore and analyze their data. We present our characterization of the petrography domain including detailed description of the experts' processes, challenges and needs. We also detail the design and development, *PetroVis*, a visualization system to support visual exploration of petrographic data. We adapt and extend the technique of parallel coordinates to simplify the analysis of the high dimensional petrographic data. We present different threads of evaluations regarding our tool including usability features and the found insight while performing a real task. Finally, we discuss some reflections and lessons we have learned and suggestions for future improvements. In the next chapter, we reflect about our experiences of working with domain experts and building visual analytic prototypes. We conclude with a set of guidelines and heuristics which would be helpful for building problem-driven visualizations for high dimensional spaces.

Chapter 6

Heuristic for Designing Interactive Visual Analytic Tools for High-dimensional Oil and Gas Data

The oil and gas domain, like other professional domains that require specific expertise and skills, has its terminology, jargon, skill-sets, challenges and needs. In particular, and as we detailed earlier in this thesis, high-dimensionality of data within the oil and gas domain is still a major challenge. Domain experts extensively use computational tools in order to simplify the challenges they face. However, many of the existing computational tools are either too general, or too specific being developed in-house for a narrow need. Most of the existing computational tools are also very expensive. Generally speaking, the existing tools do not fully satisfy the oil-and-gas industry's vision of how to efficiently pursue high-dimensional data analysis.

Computational tool designers who strive to develop new intuitive tools to satisfy the needs of the oil and gas domain experts have to start their design from scratch with little guidance since there is not general set of design guidelines to be followed.

In this chapter, we present our attempt to provide a set of guidelines to support visualization and interaction researchers, who are designing new interactive visualization analytical tools dealing with oil and gas high-dimensional data. The design heuristics presented below emerged from lessons we have learned during our exploration of the high dimensional oil gas data, including our efforts of designing, developing and evaluating of *FractVis* (discussed in Chapter 3), *Proxemic FractVis* (discussed in Chapter 4), and PetroVis (discussed in Chapter 5). We are not claiming a comprehensive compilation of strict rules that must be followed, rather a set of suggestions that we believe may shed the light on particular concerns we found common, among the high dimensional oil and gas instances that we have studied. We hope that our heuristics will be valuable for future efforts in designing interactive visual exploratory tools of high dimensional oil and gas data. Our design heuristics are:

- H1. *Empower oil and gas experts by adapting familiar domain visual representation*;
- H2. *Adapt domain terminology and mnemonics for effective communication*;
- H3. *Clarify your needs and expectations for the domain experts*;
- H4. *When collaborating on your design, match expertise with needs*;
- H5. *Facilitate correlations discovery among reservoir properties.*

Many similarities exist between our set of interactive visual analytic design heuristics, for high-dimensional oil and gas data, to existing, general classic design heuristics (e.g. [69]), or more specific guidelines for the design of problem-driven visualization systems (e.g. Sedlmair et al.. [83]). We believe that our heuristics are not negating any of these guidelines for good design, but simply add a narrower lens focusing on the specific domain of oil and gas with its particular attributes, tasks and practices.

It is worth noting that some of our design heuristics are only valid during the early phase of system development (e.g., during the phase of requirement-gathering). A primary goal of our heuristics is to improve collaboration between researchers and domain experts, and to reflect on some of the unique properties of creating such design collaboration with oil and gas experts.

The reminder of the chapter includes a description of each of our heuristics followed by an evaluation of them: considering each of them using the presented prototypes in this thesis; we do this as we aim to validate the effectiveness of the proposed heuristics and reflect on the design of our prototypes.

6.1 Heuristics

H1. Empower oil and gas experts by adapting familiar domain visual representation

Expert users in general, and oil and gas experts in particular often resist learning and using new tools, thus presenting a unique challenge for interaction designers. Although some of the oil and gas analytic-tasks are still manual and cumbersome and can clearly benefit from automation, experts regularly lack enthusiasm and complain on the difficulties of using new computational and visualization tools that could simplify their tasks. We believe that one way to encourage experts to use the new tools is by simplifying the visualization and interaction through the integration of familiar domain specific visualization elements (e.g. cross-plotting). In addition, it may be useful to adapt techniques (and inspiration) from other domains which are well known to the common users, even if these are not specific to the oil and gas domain (e.g. gamification elements [32]) while designing a new visualization for the oil and gas experts.

We believe that adapting familiar and well-known visualization could be one way to simplify the challenge of experts' resistance. Interaction designers should try to adapt the new visualization by building over and extending the ones currently used by domain experts. Such adaptation may also include embedding new visuals within the existing ones such as visualizing scatterplots within the visualization of parallel coordinates. While such integration might complicate the visualization to some extent, we believe it will help experts learn the new visualization in faster, by observing how the newly introduced visualizations are behaving while being able to quickly associate them to the existing familiar ones. For instance, microseismic engineers are trained in using Matlab for the last several decades and are familiar with visualizations, such as cross-plotting, when correlating their data. Following, integrating similar scatter plots into new microseismic visualizations can help microseismic experts learn and adopt the new visualizations quickly.

H2. Adapt domain terminology and mnemonics for effective communication

The domain of oil and gas relates to a medium and processes that are to a large extent foreign to people untrained in the specifics of the domain. Much of oil and gas processes are taking place under the ground, with data that is acquired from the surface with inherent uncertainty. Creating an effective collaboration between oil and gas domain experts and interaction designers can be very challenging due to lack of shared common ground relating to the oil and gas data and its usage.

Such difficulties are being intensified because of lack of common terminology, jargon, and possible misuse of common keywords during the communication between domain experts and interaction researchers. “Characterization”, for instance, is one example of a keyword that is very common in both oil & gas as well as HCI domain, but with completely different meanings. HCI researchers generally consider characterization as a process which involves understanding some domain by detailing its abstractions, challenges and needs. Oil and gas experts, on the other hand, link it with the geological behaviour of the reservoir. Therefore, it is crucial to develop a common language and make sure that it conveys simple and correct meanings for all the stakeholders involved in the design process to insure proper understanding and effective future communication.

We argue that it is crucial that interaction designers spend enough time learning and understanding domain processes and needs before starting the actual development. In fact, multiple meetings along with the use of investigation techniques such as contextual inquiry [44] and (rapid) ethnography [66] would be very helpful and can improve the chances of ending up with a design that will be considered effective by the oil and gas experts. Designers should make use of easily accessible, familiar and less committing mediums to communicate their early design ideas to the domain experts. One possible way to facilitate the communication of domain ideas and concepts is through the use of sketching. Some of these sketches may be useful later during the analysis process, and might suggest different representations for visualizing the oil and gas data elements. For example, a sketch showing the microseismic monitoring process would be very helpful while thinking about the possible representations to visualize the 3D microseismic events. Figure 3.15 shows an example of a sketch highlighting possible representation and interaction with a subset of the 3D microseismic events.

H3. Clarify your needs and expectations for the domain experts

Designing with oil and gas collaborators is challenging due to the fundamental difference between the two disciplines . It is extremely important to communicate clearly the design goals, the design processes and plan, and to match expectations to a realistic view of these goals. Failure to communicate these may result in the expert losing interest, finding less time and resources for a design effort that may seem less promising than expected, or even to be less accepting and resist the new visualizations and interaction techniques being developed. To alleviate these issues, it is necessary to build and maintain a mutual common ground between the designers and the oil and experts continuously and openly discussing and agreeing on the design goals, processes and timeline, throughout the collaboration.

We suggest that interaction designers should identify and clarify their needs and expectations to the domain collaborators. As an interaction designer, let your domain collaborators know your plans, design goals, design processes, schedule, interests, and ask them to be open minded when trying the new ideas that you are prototyping. For example, if you are working with microseismic engineers, make sure to establish frequent meetings with them, so they can provide you with feedback about the progress of your development as well as the implemented features. In essence, the collaboration needs to be continuous, and over a long period, with high awareness of all stakeholders of the entire process. Each side of the collaboration should agree upon clear expectations from the process to ensure effectiveness and to improve the overall design outcome.

H4. When collaborating on your design, match expertise with needs

For the designer, collaboration with good oil and gas domain experts is both crucial and can be challenging. In general, the ideal domain collaborator is an expert who has previous experience with different visualization tools, and who can convey the domain terminologies to the interaction designer. During the collaboration, the designer may be involved with junior and senior level oil and gas experts, who may have completely different type of expertise and skills, such as reservoir engineers, geologists, or drilling engineers. Having such diversity may be useful to gain overall understanding about the different disciplines, perspectives, and how they inter-correlate within the domain and the design. However, this cannot replace the need for the designer to narrow down on his/her design goals and to match them with a specific set of oil and gas domain expertise and skills.

Before designing the visualization tool, it is important for the designer to focus the collaboration with experts who are more directly related to the visualization tool specific set of tasks, requirements, and needs. For instance, if the designer aims to build an interactive visualization of reservoir models, then the collaboration should be focused on working closely with reservoir engineers. Working with other domain experts who may be less related to the design task at hand (for example, a flow simulation engineer, relating to the previous scenario) but are more accessible may be tempting but is generally not advisable. The sub-disciplines of the oil and gas domain are quite diverse and keeping the design closely in tune with the specific set of domain tasks, processes and skills, through direct collaboration with the specific domain expert is, based on our experience, crucial to the success of the design.

H5. Facilitate correlations discovery among reservoir properties

The analysis of many oil and gas tasks involves the use of statistical techniques to facilitate data correlation between varied and different facets and attributes of the geological medium. Many domain experts would like to identify the relationships between some of their data attributes such as finding the extremum values and identifying possible outliers. However, many high dimensional visualizations do not include such statistical features, thus they are limited in supporting the experts while analyzing their data. For instance, they do not provide an intuitive way to highlight the major trends and outliers within the complex data.

Interaction designers should embed and integrate common statistical methods (e.g., variance) into the developed visualization, and provide intuitive access to these methods, in order to facilitate the visual analysis of the data. Another derived concept is to consider design elements which are sensitive to statistical trends and outliers in the data and highlight them for experts. The interface can also benefit from discovering trends within the data, associating them to importance level, and representing these levels of importance visually to the user. For instance, and as discussed in the thesis, microseismic experts give more emphasis to certain independent attributes such as time and magnitude during many analytical tasks. Visualization researchers may consider coloring such attributes according to their importance to represent and highlight them in a different way from the dependent (less important) ones.

6.2 Evaluation

In this section, we use our prototypes, FractVis, Proxemic FractVis and PetroVis, as well as the processes and experience we followed and gained when designing and developing them, to evaluate, examine and demonstrate the validity of our heuristics:

- H1. *Empower oil and gas experts by adapting familiar domain visual representation*;
- H2. *Adapt domain terminology and mnemonics for effective communication*;
- H3. *Clarify your needs and expectations for the domain experts*;
- H4. *When collaborating on your design, match expertise with needs*;
- H5. *Facilitate correlations discovery among reservoir properties*.

Our aim is to reflect on the validity of the heuristics and allow readers to gain insight that could be useful to guide their own design of interactive visualization systems involving high dimensional oil and gas data. (in our discussion below we will refer to each heuristic using its corresponding code whenever needed, e.g. H2).

6.2.1 FractVis

Our initial collaboration with the microseismic experts prior to the development of *FractVis* (Chapter 3) was not based on clear expectations or common grounds (failing H3). Therefore, early prototypes of *FractVis* were misguided and ended up being discontinued. Later, we built a common language with our domain collaborators by learning the basics of the domain terminology, thus we achieved better communication (H2).

In *FractVis*, we supported interactive integration of existing familiar oil and gas visualizations such as scatterplot with the use of novel visualization techniques such as parallel coordinates (PC), aiming to improve the user experience (H1). In fact, there was strong opposition to parallel coordinates at the early phases of *FractVis*. Our expert collaborators found the new visualization difficult to grasp, and resisted learning and accepting it. As a result, we simplified our visualization and adapted it with familiar domain representations through the intuitive integration of the scatterplot visualization (H1). Later, our expert collaborators acknowledged the new visualization and highlighted its potential in helping them correlate the multidimensional microseismic data.

We believe that *FractVis* excels in providing new ways of supporting microseismic experts as they are exploring their data (H5). For instance, the support of color-correlation (Chapter 3) facilitates outliers identification and attributes correlation without the need to perform axis reordering. On the other hand, *FractVis* lacks a proper way of exploring time-related aspects of the microseismic data, in a comparison to some of the existing commercial tools. Future work regarding improving *FractVis* should provide intuitive time-analysis features and enhanced interactions to further support the microseismic experts (H1).

6.2.2 PetroVis

PetroVis was designed and developed to facilitate petrographic data analysis, following a characterization (H3) of the petrography domain. During the design of *PetroVis*, we collaborated with our domain expert while being inspired by observational techniques such as contextual inquiries and rapid ethnography in order learn more about the domain of petrography (H2). We had our domain collaborator available on-site in our research lab during the design process, so it was easier for us to continually meet to discuss and refine our design ideas. Having that enabled us to build common expectations and simplified the collaboration and the design (H3). Although we had only a narrow, well-focused collaboration with a single domain expert, we decided not to expand our collaboration by including other experts who are less related, in order to ensure that our design process is focused and in tune with the specific set of petrographic domain tasks (H4). This collaboration was very useful and fruitful for both sides, and enabled us as interaction designers to better understand the domain of petrography and its jargon.

In *PetroVis*, we adapted the visualization of scatter plot in coordination with other system visuals. However, *PetroVis* did not extend the scatter plot with additional features (failing H1) since our focus was to support high-dimensional petrographic visualization and analysis according to the needs of the domain experts (H3). Later during the evaluation of *PetroVis*, some geologists and reservoir-engineering experts highlighted the need to extend the scatter plot with statistical features to enable better correlation discovery. Although we supported correlation discovery by integrating statistical features with the parallel coordinates, we still think that further improvements are needed to extend the developed prototype and make it more useful (H5).

6.3 Summary

In this chapter, we presented a set of design heuristics with the hope of guiding future efforts of designing and developing interactive visualization systems for high dimensional oil and gas data. In addition, we showed how our heuristics can be used in practice by reflecting and evaluating them based on our experiences of designing the prototypes presented in this thesis.

Our heuristics should not be applied as a check-list but rather as a set of suggestions to help guide interaction designers and visualization researchers working with high dimensional oil and gas data. We expect that researchers and interaction designers consider and possibly apply some of the presented heuristics according to their own specific cases and design needs. At the end, the presented heuristics are meant to be used within the scope of designing for the oil and gas domain, since they reflect and are grounded on aspects which are closely related to working with the oil and gas experts and tasks.

Chapter 7

Conclusion and Future Work

In this thesis we presented our research on applying visualization and interaction techniques to enable exploration of high dimensional oil and gas data. A potential for exploring high-dimensional oil and gas data have been shown through the design, development and evaluation of three prototypes, and it is supported by insights we gained from the experts' feedback. The first prototype, *FractVis*, is a visualization system for interactive exploration of microseismic data. The second prototype, *Proxemic FractVis*, is an experimental prototype aiming to improve navigation and interaction with the 3D microseismic data using proxemics and 3D interaction techniques. Finally, the third prototype we designed, *PetroVis*, is a visualization system developed to support visual analysis of the complex petrographic data. Our prototypes presented novel applied elements by integrating and extending existing visualization and interaction techniques that are new to the current tools and workflow used in the oil and gas domain. We also presented and discussed a set of design heuristics concluding our research efforts with perspectives of relevant suggestions for future improvement. Our heuristics provided guidelines and perspectives on how to approach similar efforts of introducing new interactive visualization tools to the domain of oil and gas. To summarize, the contributions of this thesis were:

1. **Design, prototyping, implementation and evaluation of *FractVis* - a novel interactive visual analysis and exploration tool for high dimensional microseismic monitoring data**.

2. **Design and implementation of *Proxemic FractVis* – a novel interactive prototype exploring the application of proxemic interaction and 3D interaction techniques in the domain of microseismic monitoring data analysis**.

3. **Design, prototyping, implementation and thorough evaluation of *PetroVis* - a novel interactive visual analysis and exploration tool for high dimensional petrographic data**.

4. **A set of design heuristics for future design efforts in the domain of interactive visualizations of high-dimensional oil and gas data**.

7.1 Future Work

Our previous discussion of the design, implementation and evaluation of each of the prototypes highlighted suggestions and ideas for future directions (see Sections 3.8, 4.7, and 5.7. Here we present a more general broader future work discussion related to the elements shared among all our presented prototypes, namely, visual analytics (Section 7.1.1) and Ethnographic and Formal Evaluation (Section 7.1.2).

7.1.1 Visual Analytics

The field of visual analytics is a growing field with big potential for simplifying the analysis of continuously growing datasets [51]. Our developed visualization systems are merely exploratory prototypes, and our ultimate goal is to provide a complete visual analytic system for simplifying the analysis of the complex oil and gas data. However, it is still unclear how to build a comprehensive visual analytic solution for oil and gas data, due to many challenges such as the multidisciplinary aspect of the domain and the high-dimensionality of the data. Generally, noise and uncertainty are associated with the raw oil-and-gas datasets. Thus, preprocessing of the collected data is often needed before starting the actual data analysis. Future visual analytic systems should be designed to support real-time analysis if possible; that is, they should allow experts to visually analyze the data while it is being collected. Such capabilities would accelerate the exploration of the data and lead to faster informed decision-making.

A crucial question to be answered is how to integrate intuitively the oil and gas multi-scale separated datasets in a unified way. One idea toward solving that question could include examining existing computational tools focused on certain common operations, as well as trying to merge data from the various disciplines, such as geology, geophysics, and economics, onto a single entity.

We think that further research towards improving this work should also consider continuous and deeper collaboration between visualization and interaction researchers and the domain experts, and seeking a holistic representation unifying the various datasets relating to the oil and gas problem domain.

7.1.2 Ethnographic and Formal Evaluation

We developed and evaluated our visualization prototypes continuously guided by our domain collaborators. However, within the time and limitations of our work, we were not successful in conducting a full formal evaluation study with domain experts. Therefore, we think that future efforts are called for conducting a detailed evaluation study, thoroughly assessing the validity of the presented techniques in a more comprehensive way (e.g., task-oriented scenario).

In spite of the fact that we took inspiration from the method of contextual inquiry as a quick mean to learn more about the studied domain instances, we think that a comprehensive ethnographic observation should be employed to deeply understand the domain environment and other pertaining details. Generally, future efforts aiming to extend this work could benefit from field observational techniques, and perhaps would enable better understanding of experts' needs.

7.2 Final Words

In this thesis, we presented our experimental research through a set of prototypes designed and developed to enable interactive visual exploration of high dimensional oil and gas datasets. Our prototypes combined with novel techniques showed the potential benefits of employing interactive visual analytic techniques when exploring high dimensional oil and gas data. The outcome of the conducted evaluations reflected on how to further pursue this endeavor in the future. We also presented a set of design guidelines reflecting on different aspects of our work, and providing a concise set of suggestions aiming to support future research involving exploration of high dimensional oil and gas data. While there is a wide scope of future efforts to explore different forms of interactions and visualize the complex domain data, we hope that our research will shed new light and inspire future research directions leading to better solutions in this important domain.

Appendix A

Support Vector Machines for Classifying Petrographic Data from Thin-sections

Characterization of oil reservoir is very important to optimize the hydrocarbon production. However, building reservoir models is a complicated process which inherits high uncertainty. Therefore, oil and gas experts are continuously demanding better computational tools to help them understand and explore their complex data. In this appendix we outline our effort of implementing support vector machines for classifying petrographic data from Thin-sections. This effort is directly related to the PetroVis project (in Chapter 5) and has the following elements(... theoretical discussion, petrographic data classification, Matlab implementation, etc...)

A.1 Overview

Oil and gas experts highlight the importance of two reservoir properties which greatly affect the modeling of the reservoir, namely, porosity and permeability. Such properties describe clearly the distribution of the void spaces within the underground reservoir, and enable better understanding regarding how oil is flowing inside the reservoir. Consequently, they are trying to understand and estimate the behavior of these properties as well as any correlation of them with any other properties.

Existing methods from statistical learning theory (such as back-propagation neural networks) have been applied to explore building reservoir predictive models. However, most of these methods focus on estimating the empirical risk so they perform weakly on unseen data (e.g. El-sheikh and Syam [34]). Recently, some research started to adapt the use of Support Vector Machine SVMs (Vapnik 1995 [91]) for classification by taking the advantage of structural risk minimization principle (Kecman 2005 [50]), thus showing superior performance with training and unseen data. For instance, ”AL-Anazi and Gates” ([9]) developed a framework based on SVMs to identify lithology from well logs; they compared the results of their classification with results from other methods such as discriminant analysis, and revealed that SVM classifier provided better performance. Similarly, they recently published a support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs (AL-Anazi and Gates [10]).

In our approach, we explore the use of SVMs to classify petrographic data from thin-sections. Our goal is to re-classify the samples according to different classification scheme in order to better understand the porosity correlation. Our process is followed with continuous guidance and feedback from our domain collaborators. We picked five important features that we considered while classifying the data, and our goal is to identify which of them is the most important feature that affect (indirectly) the porosity values. While discussing our approach, we explain the use of cross validation over our data and the use of different kernel functions as part of our attempt to better optimize the learning model. Finally, we conclude with a discussion of the results and highlight the found insight with a comparison of the results of a visual analytic tool developed for exploring petrographic data.

A.2 Existing Classification Techniques

Techniques for data classification have been used in many domains, and the working principle of them differs according to many factors including how they estimate the learning error. Today SVMs shows better results than NNs and other statistical models, for many problems ([26, 27]). Classical NNs employs an appropriate structure of modeling the problem, and then it tries to minimize the training error (i.e. empirical risk) by fixing the estimation error (i.e. confidence interval). On the other hand, SVMs employs a different strategy by keeping the value of the training error fixed and minimize the confidence interval. Ideally, both methods have the goal of matching learning machine capacity with training data complexity. The difference of the two models comes from the minimization of different cost functions. Table A.2 shows the basic risk functionals applied to in developing the statistical models of NNs, and SVMs.

Multilayer Perceptron NN	Support Vector Machine
$R = \sum_{i=1}^{l} \left(d_i - f(x_i, w)\right)^2$	$R = \sum_{i=1}^{l} L\epsilon + \Omega\left(l, h\right)$ $L\epsilon = \lvert y - f(x, w)\rvert_{\epsilon}$

Table 1. Risk funsctionals applied for NNs and SVMs models (Kecman 2005 [50])

An interesting property about SVMs is that it creates a model with minimized VC dimension and when the VC dimension of the model is low, the expected probability of error is low as well. This means good performance on previously unseen data, i.e. good generalization. In addition, the model that generalizes well is a good model and not the model that performs well on training data pairs. This is the next important difference between the NNs and SVMs and follows from the implementation of SRM in designing SVMs, instead of a minimization of the sum of error squares, which is a standard cost function for NNs (Kecman 2005 [50]).

A.3 Support Vector Machines

SVMs is a supervised learning (from examples) technique, and it assumes no information about the underlying probability function, thus one must perform probability-free learning. In general, the basic idea behind the magic of SVM is that it minimizes an upper bound of the generalization error through maximizing the margin between the separating hyper-plane and the data (Figure A.1). The only information available is a training data set

$$D = \{(x_i, y_i)\, \epsilon X \times Y\}, i = 1, l \tag{A.1}$$

where l represents the number of the data pairs and therefore the size of the training set.

SVMs tries to classify input data classes by separating them linearly, in the input space, if possible. If the data is not linearly separable, then the SV machine first transform the input into a

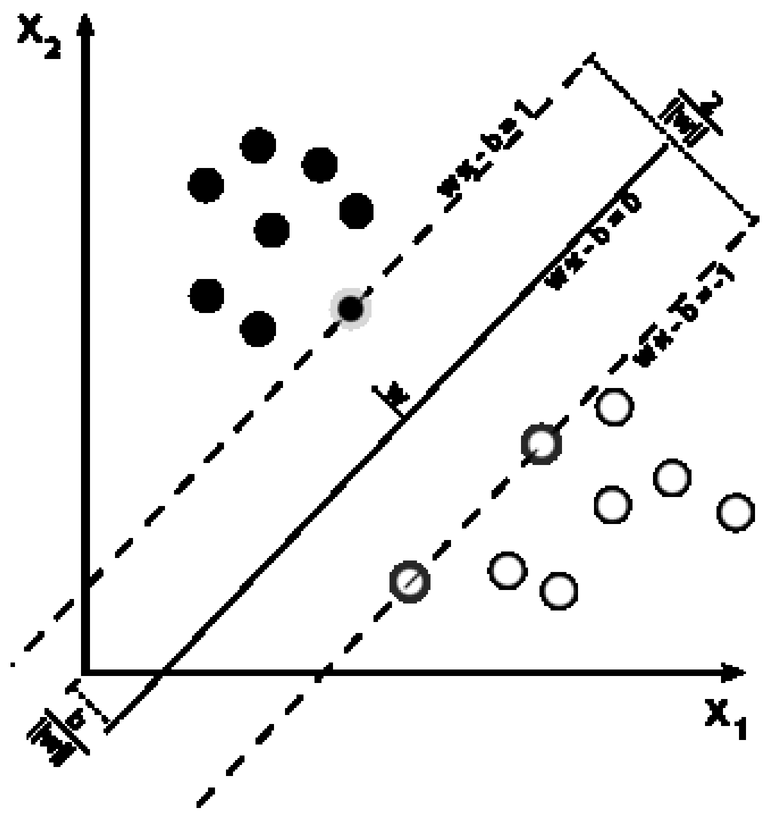

Figure A.1: Maximum-margin hyperplane and margins for an SVM trained with samples from two classes .

higher dimensional feature space and perform the separation there. The transformation can happen though a nonlinear mapping such as polynomial. The basic idea of separation is to try to find a hyperplane that isolate each group of the input samples according to the desired output. The result of the learning is an approximation function which is also called a hypothesis function. This function approximates the decision boundary, i.e., separation function, in the case of classification. The chosen hypothesis f(x,w) belongs to a hypothesis space of functions $H(f_a \in H)$ and it is a function that minimizes some risk functional R(w). A learning machine tries to capture an unknown target function f(x) that is believed to belong to some target space T, or to a class T, that is also called a concept class.

The learning machine often could provide different possible hyperplanes to separate the data, but some of these planes may overfit the data. Models that overfit the data will definitely perform badly on, during the training unseen, and the test examples. Therefore, it is better to try to separate the data using a hyperplane that maximize the margin between the data group.

The SRM is a novel inductive principle for learning from finite training data sets. It proved to be very useful when dealing with small samples. The basic idea of the SRM is to choose (from a large number of possibly candidate learning machines), a model of the right capacity to describe the given training data pairs. This can be done by restricting the hypothesis space H of approximating functions and simultaneously by controlling their flexibility (complexity).

SRM principle and its algorithmic realization through the SV machine provide the flexibility to control the separation through different parameters (Kecman 2005 [50]). Generally, The Structural Risk Minimization principle tries to minimize an expected risk (the cost function) R comprising two terms as given in Table A.2 for the SVMs $R = \Omega(l,h) + {}_{i=1}^{l} L_\epsilon = \Omega(l,h) + R_{emp}$ and it is based on the fact that for the classification learning problem with a probability of at least $1-\eta$ the bound

$$R(w_n) \leq \Omega\left(\frac{h}{l}, \frac{ln(\eta)}{l}\right) + R_{emp}(W_n), \qquad \text{(A.2)}$$

holds. The first term on the right hand side is named a VC confidence that is defined as

$$\Omega \left(\frac{h}{l}, \frac{ln(\eta)}{l} \right) = \sqrt{\frac{h\left[ln(\frac{2l}{h}) + 1\right] - ln(\frac{\eta}{4})}{l}} \tag{A.3}$$

The parameter h is called the VC dimension of a set of functions. It describes the capacity of a set of functions implemented in a learning machine (Kecman 2005 [50]). For binary classification h is the maximal number of points which can be separated (shattered) into two classes in all possible 2^h ways by using the functions of the learning machine.

Equation A.2 shows that when the number of training data increases, i.e., for $l \to \infty$, an expected (true) risk $R(w_n)$ is very close to empirical risk $R_{emp}(w_n)$ because $\Omega \to 0$. On the other hand, when the probability $1 - \eta$ approaches 1, the generalization bound grows large, because in the case when $\eta \to 0$, the value of $\Omega \to \infty$. This has an obvious intuitive interpretation (Cherkassky 1998 [26]) in that any learning machine (model, estimates) obtained from a finite number of training data cannot have an arbitrarily high confidence level. There is always a trade-off between the accuracy provided by bounds and the degree of confidence (in these bounds).

A.3.1 Linear Classification

Classifying linearly separable data is considered a simple classification. In fact, data is linearly separable if the data are correctly labeled and represented as a linear combination of the basic vector in the original space. Let's examine the structure of data that can be used for training the SV machine. Such data can be represented as a set of training points x_i belongs to R^n labeled with y_i belongs to 1,-1. Each data point can be represented by a vector of n dimension. The optimum hyperplane that separate the data is the one with the lowest training error (or the maximum separation margin). During the training, the decision function (the hyperplane) is given by: $D(x, w, b) = w^T + b = \sum_{i=1}^{n} w_i x_i + b$ where x, w belongs to R^n, b is a scalar bias, w is a weight vector, and n is the number of training points. The difficult part of finding the optimum separation function is that we only have the training data and we have no idea about the underlying probability distribution. In other words, we need to find a hyperplane with the largest margin among all hyperplanes that minimize the training error (i.e. empirical risk). A classifier with smaller margin will have higher expected risk (for unseen data). The data can classified according to a decision rule (or indicator function) as the following: $i_F = sign(d(x_p, w, b))$ for some unseen data x_p. In other words, if $d(x_p, w, b) > 0$ then x_p belongs to class 1 (y = +1), and if $d(x_p, w, b) < 0$ then x_p belongs to class 2 (y = -1).

When the values of the functions d and i_F are the same and equal to $|1|$ for the support vectors then the found hyperplane is called a canonical hyperplane, if and only if $|d| > |i_F|$ holds for all other training points at the same time (Kecman 2005 [50]). In fact, optimum hyperplanes that are selected to correctly classify the data are called canonical hyperplanes and each is constrained (defined) as the following:

$$\underbrace{min}_{x_i \in X} |w^T x_i + b| = 1 \tag{A.4}$$

The ultimate learning goal in statistical learning theory underlying SV machines is to find a canonical hyperplane that separate with a maximal margin, and it is called the optimum canonical hyperplane. The reason for having a hyperplane with a maximal margin from a limited training dataset because it will probably better classify the new data. Furthermore, asking that our hyperplane to be canonical because it will simplify finding/calculating the support vectors.

Characterization of the margin distance between the hyperplane and each class/group of data can be expressed as $M = \frac{2}{||w||}$, and this is a very interesting result. This result shows that we can maximize the margin distance by minimizing the hyperplane normal weight vector $||w|| = \sqrt{(w^T w)} = \sqrt{w_1^2 + w_2^2 + ... + w_n^2}$. Clearly, in the case of linearly separable classes empirical error equals zero ($R_{emp} = 0$ in (Eq. A.2)) and minimization of $w^T w$ corresponds to a minimization of a confidence term Ω. The optimum hyperplane specifies support vectors, i.e., training data points closest to it, which satisfy $y_j[w^T x_j + b] = 1, j = 1, N_{sv}$. For all the other (non-SVs data points) the optimum hyperplane satisfies inequalities $y_i[w^T x_i + b] > 1$. In other words, for all the data, the optimum hyperplane should satisfy the following constraints

$$y_i[w^T x_i + b] \geq 1, i = 1, ..., l \tag{A.5}$$

where l denotes a number of training data points, and N_{SV} stands for a number of SVs. Therefore the problem can be expressed as

$$minimize \quad \frac{1}{2}w^T w, \tag{A.6}$$

subject to the constraint equation A.5. This is a classic quadratic optimization problem with inequality constraints, and we can use Lagrange to solve it. In fact, we can express our problem using Lagrange as

$$L(w, b, \alpha) = \frac{1}{2}w^T w - \sum_{i=1}^{l} \alpha_i \left\{ y_i \quad w^T x_i + b \quad - 1 \right\}, \tag{A.7}$$

where α_i represent Lagrange multipliers, and ($\alpha_i \geq 0$). In other words, using these multipliers we transform each constraint (using α_i) to one degree of freedom, thus the support vectors define the hyperplane since they hold it and restrict its movement. In addition, the previous equation should be minimized with respect to w and b and should be maximized with respect to α_i. This problem can be solved in the primal space or the dual space, but we are considering the dual space approach which gives insightful results. Solving in the dual space involves the use of KKT conditions to find the optimum solution. The KKT conditions are necessary and sufficient for a maximum of equation A.7. Actually, if we have an optimum hyperplane, then w and b values represent local min values of the saddle point (w_0, b_0, α_0), and they can be verified using the gradient (derivative) of L causing the KKT conditions to be vanished. We get the following (new) constraints or KKT complementary conditions by deriving L:

$$\frac{\partial L}{\partial w_0} = 0, i.e., \quad w_0 = \sum_{i=1}^{l} \alpha_i y_i x_i, \tag{A.8}$$

$$\frac{\partial L}{\partial b_0} = 0, i.e., \quad \sum_{i=1}^{l} \alpha_i y_i = 0, \tag{A.9}$$

By substituting equations A.8 and A.9 into $L(w, b, \alpha)$ we change our dual to be $L_d(\alpha)$ as the following:

$$L_d(\alpha) = \max_{\alpha} \; -\frac{1}{2}||w||^2 + \sum_{i=1}^{l} \alpha_i \tag{A.10}$$

subject to the constraints (1) $\alpha_i \geq 0$, (2) $\sum \alpha_i y_i = 0$, and (3) $w = \sum \alpha_i y_i x_i$. Through these new constraints, and by solving L_d we are getting an optimum hyperplane that is also the optimum for $L(w, b, \alpha)$.

Interestingly, the dual equation A.10 can be rewritten as the following:

$$L_d(\alpha) = \max_{\alpha} < 1 \mid \alpha > -\frac{1}{2} \sum_{ij} \alpha_i G_{ij} \alpha_j \tag{A.11}$$

since $\sum \alpha_i =< 1 \mid \alpha >$ and $||w||^2 =< w \mid w >= \sum_{ij} \alpha_i y_i < x_i \mid \mid x_j > y_j \alpha_j = \sum_{ij} \alpha_i G_{ij} \alpha_j$. Note that the equation of L_d (Eq. A.11) is expressed in terms of the training data and depends only on the scalar products of the input patterns inside the gram matrix G. Additionally, it should be noticed that the gram matrix needs to be a positive semi-definite (PSD) matrix. Now this last expression (Eq. A.11) is a normal Lagrangian that could be solved to find α_i values, and consequently find w_0 and b_0 of the optimal hyperplane. Furthermore, we notice that the Lagrange multipliers for all non-support vectors equal zero. Finally, having calculated w_o and b_0 we obtain a decision hyperplane $d(x)$ and an indicator function $i_F = 0 = sign(d(x))$ as given below

$$D(x) = \sum_{i=1}^{l} w_0 x_i + b_0 = \sum_{i=1}^{l} y_i \alpha_i x_i^T x + b_0, \;\; i_F = 0 = sign(d(x)). \tag{A.12}$$

One final intuition about if and how linear SV machines implements the SRM principle. It can be shown that an increase in margin reduces the number of points that can be shattered i.e., the increase in margin reduces the VC dimension, and this leads to the decrease of the SVM capacity. In short, by minimizing $||w||$ (i.e., maximizing the margin) the SV machine training actually minimizes the VC dimension and consequently a generalization error (expected risk) at the same time. This is achieved by imposing a structure on the set of canonical hyperplanes and then, during the training, by choosing the one with a minimal VC dimension (Kecman 2005 [50]).

A.3.2 Soft Margin for Overlapping Classes

If the data samples are overlapped and can not be separated linearly, then it can not be solved using the previous QP method since one of the constraints can not be satisfied (Eq. A.5). Furthermore, the corresponding α_i will tends to infinity. In fact, when the algorithm finds a point which is in the 'wrong' side, it will try to increase its α_i in order to try to classify it correctly. However, even after increasing its α_i to the maximum, it can not be classified, and this affects the algorithm to use (almost) all the training points as support vectors. Thus, there should be some way to ignore that point and keep it misclassifed in the wrong side. In practice, we allow a soft margin and all data inside this margin (whether on the correct side of the separating line or on the wrong one) are neglected. The width of a soft margin can be controlled by a corresponding penalty parameter C (introduced below) that determines the trade-off between the training error and VC dimension of the model. Our problem can then be stated as the following:

$$min\ \frac{1}{2}w^T w + C\ \ (number\ of\ misclassified\ data), \tag{A.13}$$

where C is a penalty parameter, trading of the the margin size for the number of misclassified points. In other words, the corresponding α_i will not exceed C. In fact, large C leads to small number of misclassification and consequently to small margin. Clearly, taking $C = \infty$ requires that the number of misclassified data to be zero which is not possible in the case of overlapping. The possible solution considers introducing a distance measurements (represented through slack variables ξ_i) of the points crossing the margin and trade their sum for the margin size. Note that the slack variables represent the amount of violation that we should minimize. By introducing these distances as slack variables ξ_i (i=1, l), our problem can be reformulated as follows:

$$min\ \frac{1}{2}w^T w + C \sum_{i=1}^{l} \xi_i, \tag{A.14}$$

subject to

$$y_i[w^T x_i + b] \geq 1 - \xi_i, i = 1, l, \xi_i \geq 0, \tag{A.15}$$

Thus, the final quadratic optimization problem is practically the same as that for the separable case, with the only difference being in the modified bounds of the Lagrange multipliers α_i. The penalty parameter C, which is now the upper bound on α_i, is determined by the user. The selection of a "good" or "proper" C is always done experimentally by using some cross-validation technique.

Note that in the previous linearly separable case, without data overlapping, this upper bound $C = \infty$.

A.3.3 Nonlinear Classification

Some data can not be separated using linear hyperplanes, but can be separated using nonlinear decision hypersurface. Interestingly, the previous linear classifier approach can be extended to create nonlinear decision hyperplane, thus supports the ability to classify nonlinearly separable data. In fact, the way to achieve that is by doing the linear classification in a so-called feature space F. For example, imagine a set of inputs that no hyperplane can separate them within the input space, but rather they could be separated using a polynomial, for instance, without any error. The basic idea in designing nonlinear SV machines is to map input vectors $x \in \Re^n$ into vectors $\Phi(x)$ of a higher dimensional feature space F (where Φ represents mapping: $\Re^n \to \Re^f$), and to solve a linear classification problem in this feature space. We expect that this approach leads to solving a similar quadratic optimization problem with similar constraints in ϕ space. The solution for an indicator function $i_F(x) = sign(w^T\Phi(x) + b) = sign(\quad {}_{i=1}^{l} y_i\alpha_i\Phi^T(x_i)\Phi(x) + b)$, which is a linear classifier in a feature space, will create a nonlinear separating hypersurface in the original input space. Notice that the training data only appear in the form of scalar products $x_i^T x_j$. These products will be replaced by scalar products $\Phi^T(x)\Phi(x)_i$ which can be expressed using the kernel function $K(x_i, x_j)$. We also notice that the kernel function is in the input space which highlights the benefit of avoiding the need to do the mapping $\phi(x)$ at all. Instead, the required scalar product in the feature space would be calculated directly by computing kernels $K(x_i, x_j)$ for the training vectors in the input space. In addition, by applying kernels, we do not even have to know what the actual mapping $\phi(x)$ is. We can state a kernel function K as the following:

$$K(x_i, x_j) = \phi^T(x_i)\phi(x_j). \qquad \text{(A.16)}$$

In fact, there are many possible kernels that simulate the mapping, and each gives a different decision hypersurface. The choice of the kernel depends usually on the application. Table A.3.3

list some of the well known kernels.

Kernel	**Formula**
Linear (dot product)	$K(x, x_i) = (x^T x_i)$
Polynomial (of degree d)	$K(x, x_i) = [(x^T x_i) + 1]^d$
Gaussian (RBF)	$K(x, x_i) = e^{-\frac{\|\|x - x_i\|\|^2}{2\sigma^2}}$

Table 2. Common kernel functions and their mathematical expressions (Kecman 2005 [50])

Kernels are all about taking a set of features and combine (some) of them into other features and learn that. In other words, one way to think of a kernel is through the ability to pick out some features and ignore the others (by manipulating the G matrix). In fact, when we apply kernels, we recreate the Gram matrix G which will be composed of the inner-products of some vectors (in some space). The kernel will manipulate each inner-product into something else, so the Gram matrix would take the following form:

$$G = K(x_i, x_j) = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & ... & k(x_1, x_l) \\ k(x_2, x_1) & k(x_2, x_2) & ... & k(x_2, x_l) \\ ... & ... & ... & ... \\ k(x_l, x_1) & k(x_l, x_2) & ... & k(x_l, x_l) \end{bmatrix},$$

After that, we stuff our G matrix in the Lagrangian, and follow the same previous procedure for solving the problem. Finally, our decision function can be found using the following equation:

$$D(x) = \sum_{i=1}^{l} y_i \alpha_i k(x, x_i) + b. \tag{A.17}$$

A.3.4 Petrographic Data Description

Petrographic data consists of thin-sections collected at different depths from different wellbores (Figure A.2). Thin-sections are pieces of rocks prepared to study their optical properties with a petrographic microscope. In lay terms, the thin-sections represent our data samples. The gathered data are organized into a petrograhpic database. The content of the database may differ according

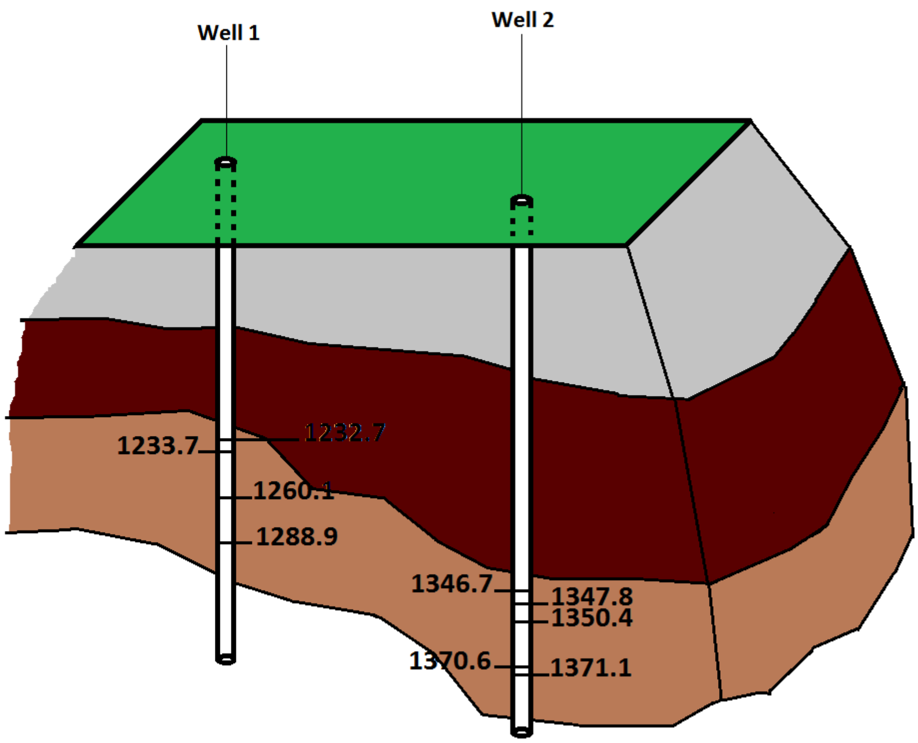

Figure A.2: Overview about the petrographic sampling from different wellbores .

to the (sparseness) of the collected samples and the accuracy of the method used in the microscopic analysis.

The scale of a petrographic database is being affected according to the number of gathered sample, the number of wells, and/or the number of identified rock properties (features). Figure A.3 shows a sample of our data. In some databases, the number of features (attributes) could be more than the number of the data samples leading to very high dimensionality of the data. Petrographic experts are demanding computational tools in order to study the complex data, understand the important features (or found minerals), and better characterize the reservoir. In fact, two important properties, permeability and porosity, greatly affect the reservoir model because they describe more about the void spaces in the rocks and the ability of oil to flow. However, the experts believe that some features (minerals) may affect indirectly the permeability or porosity values. They think that Quartz, for instance, may affect indirectly the porosity values. Therefore, classifying the data with respect to porosity would support the understanding of any correlation between the porosity and

	Well 1_1	Well 1_2	Well 1_3	Well 2_1	Well 2_2	Well 2_3	Well 2_5	...
Depths_well	1747.6	1757.4	2081.6	1232.7	1233.7	1237.8	1264.6	...
Quartz.monocrystalline	28.3	37.2	18.8	40	26.6	33.2	35	...
Quartz.polycrystalline	3.1	6	3.6	3	1	1.4	2.6	...
K.feldspar	15.3	12.8	8.8	13.6	15.5	10.5	11.4	...
Plagioclase	2.3	0.8	0	1.5	1.1	1.8	1.2	...
Biotite	1.1	1.2	2.4	1.1	1.5	5.8	5.3	...
Garnet	0.4	0.4	1.2	1.5	5.7	2.2	3.8	...
Lithics.fragments	16.8	7.6	19.6	1.1	8.6	6.1	3	...
Grain.replacive.clays	0	0	0	0	0	1.5	0	...
Kaulinite	8.8	14.8	7.6	5.7	0	0	6	...
Chlorite	0	0	0	0	0	0	0	...
Calcite	0	0	0	2.8	27	31.8	9.8	...
Dolomite	7.3	5.2	0	0	0	0	0	...
Authigenic.quartz	1.5	0.8	2	1.1	0	0.7	0	...
Pyrite	0	0.8	0.4	3	8.6	2.2	1.5	...
Others	0	0	0	0	0	0	0	...
Intergranular.porosity	8.4	5.2	1.2	19.6	0.3	0	12	...
Intragranular.porosity	4	3.2	0	1	0	1	0.4	...
Mouldic.Porosity	2.3	2.4	0.4	3	0	1.8	6.7	...
Fracture.Porosity	0.4	1.6	0	0.5	0.4	0	0	...
Oversized	0	0	0	0	0.7	0	0	...
Shrinkage	0	0	2.8	0	0	0	0.4	...
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:
CLUSTER	1	1	2	1	2	2	1	...

Figure A.3: Sample of the dataset showing its structure .

the other minerals.

In this project we used a data set which has 117 samples and 33 attributes (features). According to our domain collaborator, five features, namely: ”Moscovite”, ”Authigenic.quartz”, ”Chlorite”, ”Kaulinite”, and ”Feldspar.overgrowth” are the most possible geological attributes that could affect the porosity. Therefore, we decided to focus our classification using only those five attributes. In addition, we did not use PCA to lower the dimensionality of the data attributes, since the result of PCA may give us different important features while processing all the data features. In other words, the analysis of importance here is guided by our domain collaborators who think that it would make more geological sense to study the above mentioned five attributes. In addition, if we use all the attributes while classifying the samples, then each attribute would influence the classification process even if it does not make a geological sense to consider it at all.

A.4 SVMs to Classify Petrographic data

In this project, we explore the use of SVMs to classify petrographic data from thin-sections. Our goal is to re-classify the samples according to different classification schemes in order to identify the important features that affect the reservoir porosity. For simplicity, we consider two classes only based on **porosity**. In other words, the results would show the samples being classified as either having "*good-porosity*" or "*bad-porosity*". We chose "*porosity*" for classification because it is an important reservoir property and the porosity values of all samples are available in the data. In addition, we use each sample's values (attributes) for evaluating the sample in order to classify it. We choose the attributes that have some relationship to porosity and their values affect the porosity value. For instance, we use "Chlorite","Kaulin", "Quartz", etc. as our input vector representing each sample. We remove the "porosity" values from the table and try to classify the samples based on the other attributes and then compare the resulting classification with the actual porosity values that we already have.

Prior to performing the classification, we normalized the data in order to make sure we have homogeneous distribution of the values to prevent the dominance of one features over the others. Thus, the normalization happened for every feature by mapping the range of each feature to be between zero and one. In other words, the maximum value of any feature would be mapped to one and the minimum value would be mapped to zero. We did not use the statistical normalization approach of insuring that the mean is zero and variance is one, because we wanted to consider the full range of every attribute. On the other hand, by normalizing the data the output alphas α_i will influence each data attribute relatively leading to easier identification of the important features in our data.

Following, we describe more detail about our approach of using cross validation while building and refining our classification, and the Matlab code that we created. We conclude with a brief discussion of the results that we got.

	A	B	C	D	E	F	G
1	**Moscovite**	**Kaulinite**	**Chlorite**	**Authigenic.quartz**	**Feldspar.overgrowth**	**Porosity categories (1: high and -1:low)**	**Random Permutations**
2							
3	Fold 1						
4	0.000	0.000	0.000	0.288	0.082	-1	0.724768706
5	0.000	0.057	0.000	0.000	0.000	1	0.816381231
6	0.000	0.043	0.000	0.000	0.000	-1	0.838031874
7	0.000	0.089	0.000	0.000	0.000	1	0.274286845
8	0.000	0.000	0.960	0.416	0.265	-1	0.5682508
9	0.000	0.129	0.000	0.016	0.000	-1	0.660969019
10	0.000	0.214	0.000	0.000	0.000	1	0.004741397
11	0.125	0.243	0.000	0.032	0.082	1	0.043404667
12	0.000	0.000	0.040	0.016	0.000	1	0.297711036
13	0.000	0.000	0.000	0.048	0.122	-1	0.953085314
14	0.250	0.000	0.000	0.032	0.000	1	0.209381773
15							
16	Fold 2						
17	0.000	0.357	0.000	0.040	0.071	1	0.204574349
18	0.000	0.014	0.000	0.032	0.000	-1	0.61026552
19	0.000	0.314	0.000	0.060	0.000	1	0.497623323
20	0.094	0.125	0.000	0.028	0.031	1	0.341361106
21	0.000	0.079	0.000	0.000	0.000	-1	0.606505475
22	0.000	0.093	0.000	0.028	0.041	-1	0.509833211
23	0.125	0.186	0.000	0.000	0.000	1	0.578938111
24	0.000	0.111	0.000	0.000	0.000	-1	0.658071919
25	0.000	0.093	0.000	0.000	0.000	-1	0.304940041
26	0.000	0.118	0.000	0.028	0.000	-1	0.556171508
27	0.125	0.200	0.000	0.020	0.000	1	0.952994419
28							
29	Fold 3						
30	0.000	0.000	0.517	0.064	0.408	1	0.494562077
31	0.000	0.029	0.000	0.000	0.316	-1	0.958697415
32	0.125	0.450	0.000	0.016	0.000	-1	0.176522857
33	0.656	0.279	0.174	0.000	0.000	1	0.946202828
34	0.000	0.029	0.000	0.016	0.000	1	0.230221321
35	0.000	0.179	0.000	0.016	0.041	1	0.013183
36	0.000	0.129	0.000	0.032	0.000	-1	0.361135163
37	0.000	0.143	0.000	0.000	0.000	1	0.555623187
38	0.000	0.000	1.000	0.124	0.235	1	0.858349506
39	0.000	0.000	0.000	0.080	0.000	-1	0.297948799
40	0.219	0.000	0.761	0.028	0.000	-1	0.145235413
41							
42	Fold 4						
43	0.000	0.271	0.000	0.080	0.000	-1	0.090475267
44	0.000	0.000	0.000	0.048	0.082	-1	0.599978288
45	0.000	0.529	0.000	0.032	0.000	-1	0.497338498

Figure A.4: Petrographic data organized into subsets after random permutation of all the samples .

A.4.1 Cross-Validation

Since we have a finite amount of data, we wanted to have a better estimation of the error while classifying our data. Therefore, we decided to cross validate our data and repeat the classification randomly by reusing the same data many times. In other words, we wanted to make sure that our data has been used uniformly by trying to simulate the infinite distribution of the data. We hope that each repetition is independent enough.

First, we randomized the data by permutating all the samples once because we assumed that we do not know anything about the underlying distribution. Secondly, we divided the data into 10 subsets (folds). Each fold will be used once for testing while all remaining folds are being used for training. By repeating this process, we estimate the error of each subset/fold, and we pick that result for further analysis. In addition, while performing the learning for each fold, we tried different kernels to optimize and compare which kernel better manipulate our data. In fact, we

```
training_data = [
0.000    0.357    0.000    0.040    0.071
0.000    0.014    0.000    0.032    0.000
...
];

training_class = [
1
-1
...
];

%
% Train the SV machine
%
%SVMstruct = svmtrain(training_data,training_class,'autoscale',false,'boxconstraint',Inf,
%                      'Kernel_Function','quadratic','method','QP','showplot',false);
SVMstruct = svmtrain(training_data,training_class,'autoscale',false,'boxconstraint',Inf,
                     'Kernel_Function','linear','method','QP','showplot',false);
%SVMstruct = svmtrain(training_data,training_class,'autoscale',false,'boxconstraint',Inf,
%                      'Kernel_Function','polynomial','method','QP','showplot',false);
%SVMstruct = svmtrain(training_data,training_class,'autoscale',false,'boxconstraint',Inf,
%                      'Kernel_Function','rbf','method','QP','showplot',false);

% Compute the weight vector
dimensionsCount = 5;
W = zeros(1, dimensionsCount);
 for i=1:length(SVMstruct.SupportVectors)
     indx = SVMstruct.SupportVectorIndices(i);
     W = W + SVMstruct.Alpha(i) * training_class(indx) * SVMstruct.SupportVectors(i,:);
 end
```

Figure A.5: Matlab code to train the SV machine and calculate the weight vector .

tried the following kernels: linear, polynomial, and rbf. Figure A.4 shows an example of cross validation of the data. Detailed discussion of the results and the found insight are explained later (section A.5).

A.4.2 Implementation

We used the SVMs classification functionality of the Matlab to classify our data. We wrote a code that receives the input training data to train an SV machine. After that, we calculate the weight vector according to equation A.8. Finally, we compare the classification results with the actual measured porosity values, and calculate the misclassification rate (the estimated error).

While training, we experiment with different kernels (e.g. linear and polynomial) to obtain different classification results, in order to optimize and find the best classification model.

```
%%%%%%%%% k-fold
testing_data = [
0.000    0.000    0.000    0.288    0.082
0.000    0.057    0.000    0.000    0.000
...
];

% classify
newClasses = svmclassify(SVMstruct,testing_data,'Showplot',false);

correct_classification = [
-1
1
...
];

% compare and output the number (and rate) of misclassification
numMisClassification = 0;
for i=1:length(newClasses)
    if isequal(newClasses(i), correct_classification(i))
        numMisClassification = numMisClassification + 1;
    end
end

disp(['Misclassification Number ' num2str(numMisClassification)]);
disp(['Misclassification Rate ' num2str(numMisClassification / length(newClasses))]);

disp('done');

```

Figure A.6: Matlab code to train the SV machine and calculate the weight vector .

K-Fold	Misclassification Number	Misclassification Rate
Fold 1 using linear kernel	5	0.45455
Fold 1 using quadratic kernel	5	0.45455
Fold 1 using polynomial kernel	8	0.72727
Fold 2 using linear kernel	6	0.54545
Fold 2 using quadratic kernel	7	0.63636
Fold 2 using polynomial kernel	4	0.36364
Fold 3 using linear kernel	5	0.45455
Fold 3 using quadratic kernel	6	0.54545
Fold 3 using polynomial kernel	7	0.63636
Fold 4 using linear kernel	6	0.54545
Fold 4 using quadratic kernel	7	0.63636
Fold 4 using polynomial kernel	8	0.72727
Fold 5 using linear kernel	5	0.45455
Fold 5 using quadratic kernel	7	0.63636
Fold 5 using polynomial kernel	6	0.54545
Fold 6 using linear kernel	6	0.54545
Fold 6 using quadratic kernel	7	0.63636
Fold 6 using polynomial kernel	8	0.72727
Fold 7 using linear kernel	6	0.54545
Fold 7 using quadratic kernel	7	0.63636
Fold 7 using polynomial kernel	7	0.63636
Fold 8 using linear kernel	6	0.54545
Fold 8 using quadratic kernel	4	0.36364
Fold 8 using polynomial kernel	4	0.36364
Fold 9 using linear kernel	9	0.81818
Fold 9 using quadratic kernel	6	0.54545
Fold 9 using polynomial kernel	5	0.45455
Fold 10 using linear kernel	8	0.66667
Fold 10 using quadratic kernel	10	0.83333
Fold 10 using polynomial kernel	6	0.5

Figure A.7: Result of data classification .

Kernel type	Best fold(s)
linear	Fold1 Fold 3
quadratic	Fold 8
polynomial	Fold 2 Fold 8

Figure A.8: Folds with minimum (classification) error according to kernel type .

A.5 Results and Discussion

In this section, we present the results of performing the classification using different kernels, and for each data fold. We also compare the classification results with the results from a visualization tool developed to simplify the analysis of petrographic data. Then we highlight the insight that we could conclude by analyzing all the results.

During our experiments, we recorded the misclassification number and the misclassification rate by applying SVMs once for each fold with different kernels (Figure A.7). To analyze the results, we started by examining the error associated with each result from each fold.

We compared the misclassification number (and rate) among all the used kernels. We found that Fold 1 and Fold 3 are the best folds when using the linear kernel because the reflect the minimum error. Similarly, we found that Fold 8 is the best fold when using the quadratic kernel, and Folds 2 and 8 are the optimum when using the polynomial kernel (Figure A.8).

After that, we examined the components of the weight vectors which are being associated with the previously found best folds. The analysis involved extracting the minimum 2 sub-component values. Since each sub-component correspond to a feature, then we are highlighting the two most important features by extracting the minimum W_i. The reason is that there is a correlation between the calculated W_i and the classification margin. In particular, the better the classification is, the larger the margin is, and proportionally the smaller the weight is. Figure A.9 shows the result of extracting the minimum weights. Interestingly, we can see that the second feature has been confirmed with among all the best folds, thus, we can conclude that the second feature, namely

Fold Number (kernel)	Weight vector (w1, w2, w3, w4, w5)	Minimum W(s)
Fold 1 (linear)	-391112121.369764 -928692695.126842 -399255674.549130 -206691047.056671 -391449870.526273	w2, w3
Fold 2 (polynomial)	-864938.655468113 -142896761.144849 -228929.724870056 -42620696.2390736 -8336192.82651809	w2, w4
Fold 3 (linear)	-341185405.401776 -894605547.780186 -294551997.213491 -200667823.500156 -347402786.848921	w2, w5
Fold 8 (polynomial)	-1512349.45529763 -158474036.853939 -626859.705145068 -32934554.3275485 -14452364.6627334	w2, w4
Fold 8 (quadratic)	-157161778.446994 -436125847.165918 -119837993.482059 -94127628.9960839 -152925537.657073	w2, w1

Figure A.9: The analysis of the results showing the optimum weights associated with the best folds
.

”Kaulinite”, is the best one to classify the data with respect to porosity.

Our previous result has been confirmed when we trained our model with ALL the training vectors. Clearly, the result highlights that the second feature is the most important one (Figure A.10).

Besides applying SVMs to classify the petrographic data, which already gave interesting results, we decided to support our argumentation by comparing the results that we have got from SVMs with the graphical intuition that we can depict from a petrographic visualization tool. In fact, we used the same tool that we developed previously to enable the visual analysis of petrographic data as part of the computational methodology to characterize the data petrofacies (Cevolani et al. 2013 [24]).

The visual results (shown in Figures A.11 and A.12) reflects almost the same conclusion drawn above. Within the visual results, we can see the the second feature is possibly the most important one that affect indirectly the porosity. In particular, the correlation shown in Figure A.12 clearly highlights this insight.

Kernel Type	Weight vector (w1, w2, w3, w4, w5)	Minimum W(s)
Linear	-436093852.623485 -984753057.281544 -422844412.538641 -218508708.546888 -401240555.591324	w2, w1
Quadratic	-204486751.935462 -519495069.458000 -153273098.757752 -114607832.482669 -169340563.318314	w2, w1
Polynomial	-8407546.63536178 -207679344.534954 -441178.182822799 -54946726.7144935 -21075808.0387560	w2, w4

Figure A.10: The analysis of the results showing the optimum weights by considering all the data samples with respect to different kernels .

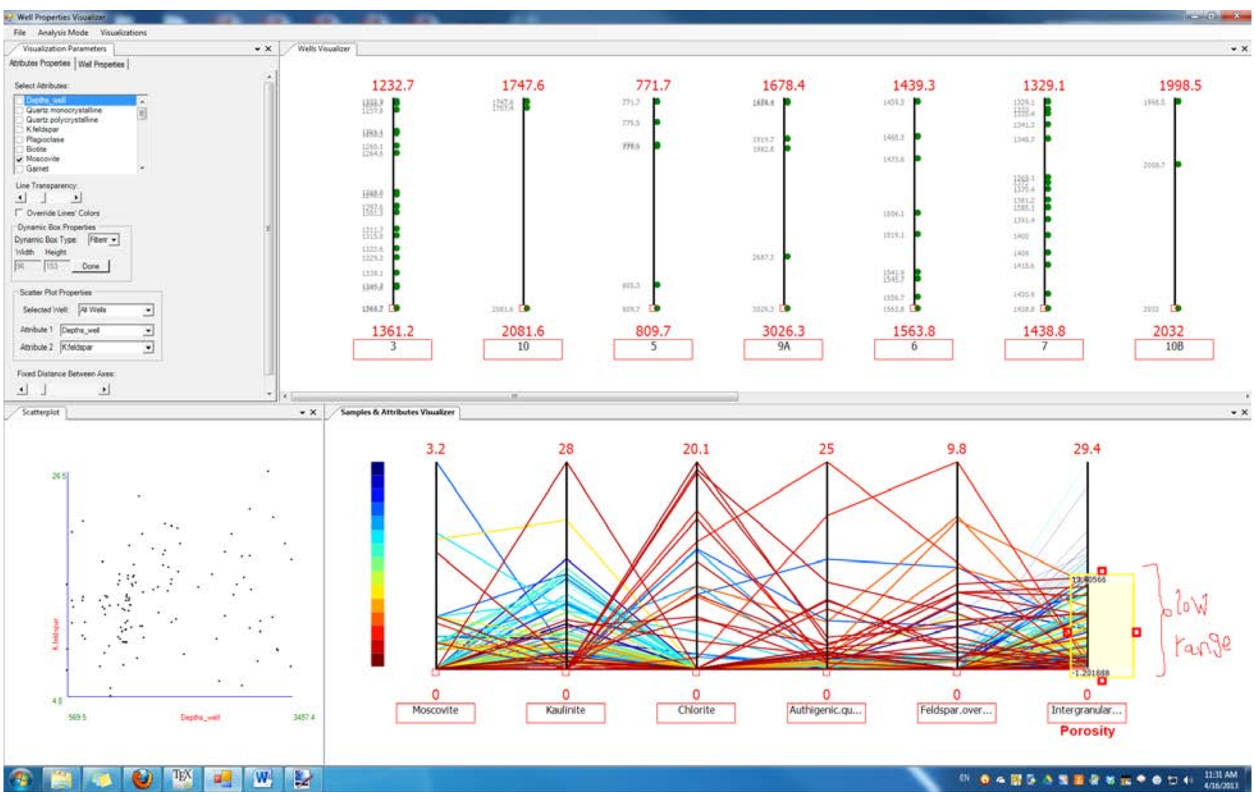

Figure A.11: The visualization does not show a clear trend (correlation) among the data features while considering low porosity range .

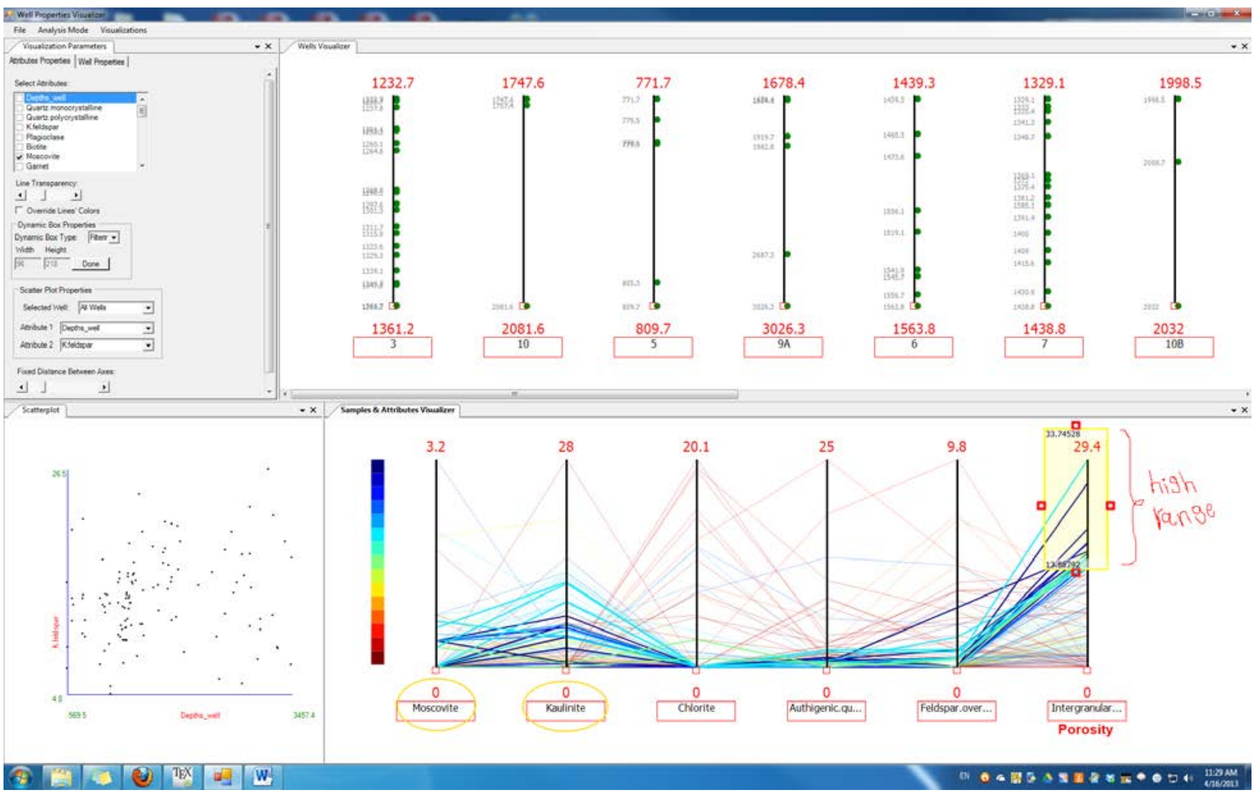

Figure A.12: The visualization highlights a clear trend (correlation) particularly for the first and second features while considering high porosity range .

Appendix B

Study and Evaluation Materials

This appendix contains the related components used for conducting user evaluations described in Chapter 3.

- description of the procedure of the study.
- study questions
- ethics approval (including approval of study modifications)
- recruitment letters used for the study.

The evaluation conducted in Chapter 5 did not follow a formal structure. On the other hand, we followed an iterative design approach with domain expert collaborators.

B.1 Evaluation Session Details

Hello! My name is Ahmed. Thank you for being part of the study today. The study that I am doing today is to evaluate different features we developed in the context of microseismic monitoring for visual exploration of the microseismic multidimensional data.

Filling the following table will provide us with basic information about you and your background. As a participant, your information will be kept anonymous and confidential. Your participation is highly appreciated and $15 will be given as reimbursement for your time and effort. This session will be video recorded for later analysis, and only with your consent.

Interview ID	
Interviewer Name	Ahmed Mostafa
Participant Name	
Participant Background/Specialization	
Date of interview	

B.2 Pre-study Questionnaire's Questions

Prior to starting the evaluation session with each participant, and after completing the study ethics clarification, we asked each participant the following questions as a pre-study questionnaire.

1. Have you ever worked with (micro) seismic data or multidimensional analysis tools before?
2. Are you familiar with (data analysis using) Scatter/Cross plot?
3. Do you have any experience in oil/gas domain (with any commercial software)?
4. Please explain which kind of results that you are expecting when you analyze your data?

5. Please describe some of the major tasks that could be important while working with your data?
6. Please describe how would you like to interact with your data?
7. Please list what do you expect from your computational tools?
8. Are you satisfied with current computational tools that you are using?
9. Would you like to see different ways (that might be more intuitive) for performing your tasks? Would you like to have better tools?
10. Do you have any experience with data filtering? Which kind of data? How important is it to have this operation in your toolset?
11. Do you think that the traditional way of supporting filtering (i.e. using sliders) is intuitive and satisfy your requirements?
12. Would you be interested to see a different way (that can be intuitive) of performing filtering?
13. Do you have any experience with data correlation? How important is it to have this operation in your toolset?
14. Do you work with multidimensional data? How do you perform the multivariate analysis among the data variables?
15. Are you mostly performing the analysis with only two variables each time? Do you think that correlating more than two attributes (finding relations, detecting outliers, etc) simultaneously is useful? Is it supported in a proper way in your toolset?
16. Would you be interested to see a different way of analyzing the attributes' correlation just by looking at a graph?

Overview about the details of the study

Let's start by giving you a basic background about the domain representing the context of our work ...

Microseismic data (MS) represents a time-varying point cloud of events happening below the earth's surface. The data is organized into different stages where each stage represents subset of the events as part of the whole microseismic data. Each event is being described using many attributes so the data is multidimensional. However, the data has lots of noise coming from the measurement's methods as well as the preprocessing. Analysis of such events and their attributes will improve the decision making process for optimizing the oil/gas production in general. Our tool has been developed to explore this data and to aid in gaining insights from it.

This study is being conducted to provide a case-study that reflects how our tool can be effective for the researchers and domain experts. We aim to validate also that our extended parallel coordinates' implementation is useful for showing and revealing insights from the data.

In short, the study goals are: (1) report the participants' ability to answer the study's questions, use our tool and find some value in it even if they do not like it; (2) report the needs if the domain participants, and how our tool can help them achieve some of these needs, and (3) report that our tool propose novel ways over our extended parallel coordinates' visualization for easier multidimensional data analysis as well as reducing the learning gab for non-visualization specialists.

This interview will flow in a conversation like-structure. Some background will be giving for clarification before some questions, if needed. You should answer the questions as required, and you can ask for clarification or more explanation if some points are not clear. For your interest, the study should take around 30-45 minutes.

Our study will be organized into 3 parts:

- **Training/Tutorial**: Give the participants proper background about the basic usage of our tool and its features and train them with simple training tasks. Some training tasks should involve the participant to interact with the tool by him/her-self directly. The training or tutorial should focus more on the things that will be tested in the next section, but minimal information only should be given in this section.

- **Task-Evaluation**: Provide the participants with some tasks and ask them to perform them with no or minimal interference from the researcher's side. Recording participant's response is very important here for later analysis.

- **Interview and Discussion**: interviewing the participants about his/her opinion of the tool, general comments and feedback.

Questions for building a semi-structured questionnaire

<u>Section A: Training/Tutorial</u>

This section will start by giving background about our tool and how to interact with it with emphasis on the features that will be tested in the next section.

Below are some questions that may be giving as training tasks to give the user a chance to interact with our tool.

The tutorial shall focus on the description of the usage of parallel coordinates (pc) in general. Then it will give specific minimal information about the following (to give background for the next section):

- Examining pc to find outliers/extremum

- The use of color mapping to quickly associate and understand the relations between the attributes

- The ability to perform axis reordering to show relation between attributes in sequence

- The ability to use lenses to filter and to show embedded visualization such as scatterplot

- The ability to use size mapping to resize the events' spheres in the 3D according to the current focused attribute, as well as a basic idea about how to navigate the 3D and understand its contents such as the wells

1. Have you ever seen or used Parallel Coordinates before? If No, then explain it with image.
2. Which microseismic stage has the lowest "magnitude" value?
3. If color mapping is not available over the parallel coordinates, Could you find the correlation between the "energy" and "moment" as non-sequenced attributes?
4. Can you select a subset of all the events from the last 3 stages?

<u>Section B: Task-Evaluation (Experimental sub-session)</u>

5. Are there any events' outliers in the microseismic "radius" attribute? [Find Anomalies] Which microseismic stage(s) has (or have) these outlier(s)? [Associate]
6. Is there a correlation between the "time-stamp" & "magnitude" for all events? You may use 'Color-mapping', 'Axis re-ordering' or 'scatter-plot lens', which method is easier (intuitive)? [Correlate]
7. Show only the events from the stage(s) that has/have the lowest "moment", "stress-drop", "energy", and the highest "magnitude"? [Identify, Filter]
8. Please categorize the microseismic stages into two groups: one that has high "magnitude" and the other that has low "magnitude". Which group has more microseismic stages? [Categorize]
9. Find a subset of events that has no outliers regarding the "energy" and "radius" attributes and update the size of the spheres of those events to be relative to the "magnitude" values distribution?

10. If you know that “MS-Distance” represents the distance between the event and the ’monitoring well (red)’ or the measuring sensor for each microseismic event, so this distance might give some idea about the quality of the event’s measurement (For example, closer events are more likely to have higher quality and less uncertainty in their measurements). Try to use this information to confirm any possible outlier(s) regarding the “SP-Moment” with or without the option of “size-mapping”. What do you think? Write down if you can confirm them using “color-mapping” only. [Open-Ended]

Section C: Discussion (post-task questionnaire)

11. Give your opinion about the importance of using only color-mapping, only size-mapping or color and size mapping together?
12. Is it useful to be able to manually interact (i.e. remove) with specific event(s) in the 3D visualization?
13. Is it useful to see different visualizations (i.e. histogram, axis scaling) embedded inside the parallel coordinates? Do you think that such embedded visualizations makes it easy to understand how parallel coordinates work and how to use it?
14. Is it useful to simulate the fracture’s growing (and events’ population) by resizing a filter box?
15. Is it useful to see the neighboring events as transparent (context) around the current selected subset (focus)?
16. Is it useful to spatially analyze two or more events to find if they are similar or not?
17. Do you think that complex filtering (i.e. filtering more than one attribute) using our filter-boxes can be more useful/intuitive than filtering with the traditional sliders?

18. Do you think that it is easy/doable/useful to analyze/correlate (using traditional methods such as scatter/cross plot) more than two attributes at the same time as in the previous question?

19. Do you think that using color mapping over any attribute may allow for revealing any correlations easily?

20. Do you think that Parallel Coordinates' technique is useful for the analysis of the microseismic attributes?

    Please choose one of the following:

Strongly Useful	Slightly Useful	I am not sure

21. If your data is time-variant, would you like to see more focus on performing time-based analysis?

22. To what degree do you think that filtering using dynamic filter boxes is useful?

    Please choose one of the following:

Strongly Agree	Slightly Agree	Neutral	Slightly Disagree	Strongly Disagree

23. Do you like Parallel Coordinates?

    Please choose one of the following:

Strongly Like	Slightly Like	Neutral	Slightly Like	Strongly Like

24. Please rank the feature of color mapping used in parallel coordinates' lines with the synchronization of the 3D events? (from 1 to 10, where 1 is the lowest)

25. Please rank the feature of size mapping over the parallel coordinates' attributes and its synchronization with other visualizations?

26. Please rank the feature of having magic-like lenses (dynamic boxes) over the parallel coordinates that could provide different visualizations inside the parallel coordinates?

27. Please provide a comment about this visualization tool in general. (I.e. what you like, what you do not like, things to be improved, etc. ...)

28. Give a comment regarding the features of this tool compared to other tools that you might have used for doing similar analysis?

Thank you very much for your time and effort

Evaluation Questionnaire

The following inquiries have been gathered through semi-structured interviews and discussion with two microseismic domain experts. The goal was to identify the needs, and tasks of the domain experts. We present the list of questions and a sample of the answers from one of the experts.

	Question	**Summarized Answer from the expert**
1	List some operations that are being complicated while being performed through your computational tools?	The inclusion of uncertainty in SRV calculation.
2	List the most important attributes in the data?	• focal mechanism • magnitude, distance, time, location ,. . .
3	List pairs of attributes that are usually analyzed together?	• depth vs. magnitude • pressure vs. event-time and location • distance vs. magnitude • injection rate vs. fracture growth • magnitude vs. geo-mechanical properties (i.e. pumping curves)

4	How to represent or visualize each event?	The default view of coloring the events should be by stage then sized (by magnitude as standard but it also) by other attributes. They said it also makes sense to color/size the events relative to other attributes but it should be intuitive mapping.
5	Do you need to filter the data by using multiple filters at the same time?	We currently do the filtering one by one, and it would be useful if we can perform multiple filtering at the same time.
6	Do you need synchronization between: the 3D visualization of events', the visualization of the attributes, and the visualization of the engineering curves?	Yes it is very important but need to be intuitive and not overwhelming.
7	Are you interested in correlating many attributes at the same time instead of correlating each two attributes together?	Yes, and we are currently doing each two one at a time.
8	How do you find and analyze data outliers?	This often happens (manually) prior to visualization and is fixed before opening/running the visualization system.

9	What kind of results you are expecting when you analyze the data?	• To obtain and understand fracture geometry and look at the fractures' interactions (i.e. pressure, stress, fracture size) relative to each other or between wells • correlation of microseismic response with well performance
10	What are the major tasks that are important in your work?	• Measurements of fracture azimuth, width, etc. • intuitive data filtering and correlation • SRV calculation • stress inversion (analysis)
11	Are you satisfied with the current tools?	Yes but it can be improved.
12	Would like to see different (better) ways for performing your tasks?	Yes, such as the ability to view the data in 3D and in 2D at the same time.
13	Do you perform any sort of data filtering and is it important to have that in your toolset?	It is critical to have this and we perform "seismic signals" filtering among other types of filtering.
14	Do you think that the use of sliders for filtering is satisfying?	Yes if it is supported with the ability to enter absolute amount.

15	Do you think that data correlation is important?	It is not critical but still important.
16	What tools do you use to perform multivariate analysis?	Excel or Matlab.
17	Would like to do this correlation just by looking at a graph?	Yes.
18	Are you interested in performing multiple attributes correlation simultaneously?	Usually, we do not do this analysis but we think it can be useful only if well-organized or have been provided with a convincing approach.

We attached the following notice during some of the assessment sessions that we had with our domain collaborators in their lab.

Notice

Observational session is in progress

Sorry for any inconvenience.

We would like to inform you that an observational session is in progress currently inside this lab/office. We appreciate, in case that you want to interact with the "Name-here" to approach me first so we could pause our recording if needed. Please find me inside the lab nearby the Name's desk.

Thank you very much for your understanding

Bibliography

[1] Endeeper petroledge for management of reservoir petrographic information. `http://www.endeeper.com/products/software/petroledge`. Accessed: 10/03/2013.

[2] ESG Solutions hydraulic fracture mapping. `https://www.esgsolutions.com/english/view.asp?x=741`. Accessed: 31/03/2012.

[3] Matlab – the language of technical computing. `http://www.mathworks.com/products/matlab`. Accessed: 07/06/2013.

[4] Petrographic analysis: how it works. `http://www.utoronto.ca/nmc/mason/petromethod.html`. Accessed: 29/01/2013.

[5] Schlumberger Petrel Solutions microseismic evaluation. `http://www.slb.com/services/software/geo`.

[6] Terraspark geoscience. `http://www.terraspark.com/microseismic/`. Accessed: 25/03/2013.

[7] Virtual spray paint. `https://blog.almostlogical.com/2008/11/09/virtual-spray-paint-part-1/`. Accessed: 23/03/2013.

[8] A. Al-Anazi and I. Gates. On the capability of support vector machines to classify lithology from well logs. *Natural Resources Research*, 19:125–139, 2010.

[9] A. Al-Anazi and I. Gates. On the capability of support vector machines to classify lithology from well logs. *Natural Resources Research*, 19(2):125–139, 2010.

[10] A. Al-Anazi and I. Gates. A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Engineering Geology*, 114(34):267 – 277, 2010.

[11] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *INFOVIS 2005.*, pages 111 – 117. IEEE Computer Society, 2005.

[12] R. Amorim, N. Boroumand, E. Vital Brazil, Y. Hajizadeh, D. Eaton, and M. Costa Sousa. Interactive sketch-based estimation of stimulated volume in unconventional reservoirs using microseismic data. In *Proceedings of 13th European Conference on the Mathematics of Oil Recovery (ECMOR XIII)*, 2012.

[13] G. Andrienko and N. Andrienko. Coordinated multiple views: a critical view. In *CMV '07*, pages 72–74, july 2007.

[14] K. W. Arthur, K. S. Booth, and C. Ware. Evaluating 3d task performance for fish tank virtual worlds. *ACM Trans. Inf. Syst.*, 11(3):239–265, July 1993.

[15] T. Ballendat, N. Marquardt, and S. Greenberg. Proxemic interaction: designing for a proximity and orientation-aware environment. In *ITS '10*, pages 121–130, New York, NY, USA, 2010. ACM.

[16] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and magic lenses: the see-through interface. In *Proc. of the 20th annual conf. on Computer graphics and interactive tech.*, SIGGRAPH '93, pages 73–80, 1993.

[17] D. Borland and R. Taylor. Rainbow color map (still) considered harmful. *Computer Graphics and Applications, IEEE*, 27(2):14–17, 2007.

[18] I. Bowman, S. Joshi, and J. Van Horn. Query-based coordinated multiple views with feature similarity space for visual analysis of mri repositories. In *Proc. of VAST' 2011*, pages 267–268, 2011.

[19] G. C. Burdea and P. Coiffet. *Virtual Reality Technology*. John Wiley and Sons, Inc., New York, NY, USA, 2 edition, 2003.

[20] J. L. Calvez, R. Klem, L. Bennett, A. Erwemi, M. Craven, J. Palacio, and Schlumberger. Real-time microseismic monitoring of hydraulic fracture treatment: A tool to improve completion and reservoir management. pages 29–31. SPE, 2007.

[21] M. Carpendale. Considering visual variables as a basis for information visualisation. Technical Report 2001-693-16, Department of Computer science, University of Calgary, Calgary, AB, Canada, 2003.

[22] M. S. T. Carpendale, D. J. Cowperthwaite, and F. D. Fracchia. Extending distortion viewing from 2d to 3d. *IEEE Comput. Graph. Appl.*, 17(4):42–51, July 1997.

[23] M. S. T. Carpendale and C. Montagnese. A framework for unifying presentation space. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, UIST '01, pages 61–70, New York, NY, USA, 2001. ACM.

[24] J. T. Cevolani, A. E. Mostafa, E. A. V. Brazil, L. C. de Oliveira, L. G. da Fonseca, and M. C. Sousa. Computational methodology to study heterogeneities in petroleum reservoirs. In *EAGE Annual Conference & Exhibition incorporating SPE Europec*, 2013.

[25] J. Chambers. *Graphical methods for data analysis*. Chapman & Hall statistics series. Wadsworth International Group, 1983.

[26] V. S. Cherkassky and F. Mulier. *Learning from Data: Concepts, Theory, and Methods*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1998.

[27] F. Chu and L. Wang. Gene expression data analysis using support vector machines. In *Neural Networks, 2003. Proceedings of the International Joint Conference on*, volume 3, pages 2268–2271 vol.3, 2003.

[28] W. C. Cleveland and M. E. McGill. *Dynamic Graphics for Statistics*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 1988.

[29] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the cave. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '93, pages 135–142, New York, NY, USA, 1993. ACM.

[30] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The cave: audio visual experience automatic virtual environment. *Commun. ACM*, 35(6):64–72, June 1992.

[31] L. F. De Ros and K. Goldberg. Reservoir petrofacies: a tool for quality characterization and prediction. In *Proceedings of AAPG Annual Convention*, 2007.

[32] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon. Gamification. using game-design elements in non-gaming contexts. In *PART 2 – Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, pages 2425–2428. ACM, 2011.

[33] D. Dopkin and H. James. Trends in visualization for e&p operations. *First Break*, 24(3), 2006.

[34] T. S. El-sheikh and M. Syiam. An efficient technique for lithology classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 27(5):629–632, 1989.

[35] G. Ellis. The sampling lens:making sense of saturated visualisations. In *Proc. CHI'05 Extended Abstracts on Human Factors in Computing Systems*, pages 1351–1354. ACM Press.

[36] N. Elmqvist, P. Dragicevic, and J.-D. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE TVCG 08*, 14(6):1539–1148.

[37] N. Elmqvist, J. Stasko, and P. Tsigas. Datameadow: A visual canvas for analysis of large-scale multivariate data. In *VAST' 2007*, pages 187–194, 30 2007-nov. 1 2007.

[38] G. W. Furnas. Generalized fisheye views. *SIGCHI Bull.*, 17(4):16–23, Apr. 1986.

[39] A. Gooch, B. Gooch, P. Shirley, and E. Cohen. A non-photorealistic lighting model for automatic technical illustration. In *Proc. of the 25th annual conf. on Computer graphics and interactive tech.*, SIGGRAPH '98, pages 447–452, 1998.

[40] E. T. Hall. *The Hidden Dimension*. Doubleday, 1996.

[41] C. Hand. A survey of 3d interaction techniques. *Computer Graphics Forum*, 16(5):269–281, 1997.

[42] T. Höllt, J. Beyer, F. Gschwantner, P. Muigg, H. Doleisch, G. Heinemann, and M. Hadwiger. Interactive seismic interpretation with piecewise global energy minimization. In *PacificVis' 2011*, pages 59–66, march 2011.

[43] D. Holten and J. J. Van Wijk. Evaluation of cluster identification performance for different pcp variants. *Computer Graphics Forum*, 29(3):793–802, 2010.

[44] K. Holtzblatt and S. Jones. *Contextual inquiry: a participatory technique for system design*, pages 177–210. Lawrence Erlbaum Associates, Hillsdale, 1993.

[45] T.-J. Hsieh, C.-K. Chen, and K.-L. Ma. Visualizing field-measured seismic data. In *PacificVis' 2010*, pages 65–72.

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

[47] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *VIS'90*, pages 361–378. IEEE Computer Society Press, 1990.

[48] I. T. Jolliffe. *Principal Component Analysis*. Springer, 2st edition, 2002.

[49] E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '01, pages 107–116, New York, NY, USA, 2001. ACM.

[50] V. Kecman. Support vector machines an introduction. In L. Wang, editor,*Support Vector Machines: Theory and Applications*, volume 177 of *Studies in Fuzziness and Soft Computing*, pages 1–47. Springer Berlin Heidelberg, 2005.

[51] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 19(3):495 –513, march 2013.

[52] D. H. Laidlaw, C. Jackson, D. Keefe, and S. Zhang. Subjective usefulness of cave and fish tank vr display systems for a scientific visualization application. In *IEEE Visualization 03 Poster Abstract*. Press, 2003.

[53] C.-H. Lee, C.-L. Liu, Y.-A. Chen, and Y.-S. Chen. Painting in the air with wii remote. *Expert Syst. Appl.*, 38(12):14668–14678, 2011.

[54] M. Lin, W. Baxter, V. Scheib, and J. Wendt. Physically based virtual painting. *Commun. ACM*, 47(8):40–47, Aug. 2004.

[55] L. F. Lu, M. L. Huang, and T.-H. Huang. A new axes re-ordering method in parallel coordinates visualization. In *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, volume 2, pages 252–257, Dec.

[56] W. E. Mackay, A.-L. Fayard, L. Frobert, and L. Médini. Reinventing the familiar: exploring an augmented reality design space for air traffic control. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pages 558–565, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.

[57] J. Marbach, B. Kadlec, and J. Carlson. Multiattribute Visualization Using Multivariate Volume Rendering and Glyphs. In *Attributes: New Views on Seismic Imaging – Their Use in Exploration and Production: 31st Annual*, volume 31, pages 140–158. SEPM - Society of Economic Paleontologists and Mineralogists, 2011.

[58] N. Marquardt. Proxemic interactions in ubiquitous computing ecologies. In *Proceedings of CHI EA '11*, pages 1033–1036, NY, USA, 2011. ACM.

[59] N. Marquardt, T. Ballendat, S. Boring, S. Greenberg, and K. Hinckley. Gradual engagement: facilitating information exchange between digital devices as a function of proximity. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, ITS '12, pages 31–40, New York, NY, USA, 2012. ACM.

[60] N. Marquardt, R. Diaz-Marino, S. Boring, and S. Greenberg. The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. In *Proceedings of UIST '11*, pages 315–326, NY, USA, 2011. ACM.

[61] N. Marquardt and S. Greenberg. Informing the design of proxemic interactions. *IEEE Pervasive Computing*, 11(2):14–23, April-June 2012. Earlier somewhat different version as Report 2011-1006-18, July 2011.

[62] A. R. Martin and M. O. Ward. High dimensional brushing for interactive exploration of multivariate data. In *VIS'95*, pages 271–278. IEEE Computer Society, 1995.

[63] K. Matkovic, M. Jelovic, J. Juric, Z. Konyha, and D. Gracanin. Interactive visual analysis and exploration of injection systems simulations. In *Visualization, 2005. VIS 05. IEEE*, pages 391–398, 2005.

[64] S. Maxwell. Microseismic: Growth born from success. *The Leading Edge*, 29(3):338–343, 2010.

[65] M. Meyer, T. Munzner, and H. Pfister. Mizbee: A multiscale synteny browser. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):897–904, 2009.

[66] D. R. Millen. Rapid ethnography: time deepening strategies for hci field research. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, DIS '00, pages 280–286, New York, NY, USA, 2000. ACM.

[67] A. Mostafa, R. Amorim, E. Vital Brazil, D. Eaton, S. Carpendale, E. Sharlin, and M. Costa Sousa. Exploratory visual modeling and analysis of microseismic events. In *Proceedings of GeoConvention 2012: Vision*, 2012.

[68] C. B. Nielsen, S. D. Jackman, I. Birol, and S. J. M. Jones. Abyss-explorer: Visualizing genome sequence assemblies. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):881–888, Nov. 2009.

[69] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 249–256, New York, NY, USA, 1990. ACM.

[70] P. Norm Warpinski. Microseismic monitoring: Inside and out. *Journal of Petroleum Tech*, 61(11):80–85, 2009.

[71] M. Novotny and H. Hauser. Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):893–900, Sept. 2006.

[72] H. Obermaier, M. I. Billen, K. I. Joy, H. Hagen, and M. Hering-Bertram. Visualization and multivariate clustering of scattered moment tensors. *Information Visualization*, 11(1):43–59, 2012.

[73] N. Ohno and A. Kageyama. Scientific visualization of geophysical simulation data by the cave vr system with volume rendering. *Physics of the Earth and Planetary Interiors*, 163(14):305 – 311, 2007. ¡ce:title¿Computational Challenges in the Earth Sciences¡/ce:title¿.

[74] S. M. Omohundro. Five balltree construction algorithms. Technical report, International Computer Science Institute, 1989.

[75] D. Patel, S. Bruckner, I. Viola, and E. Groller. Seismic volume visualization for horizon extraction. In *PacificVis' 2010*, pages 73–80, march 2010.

[76] W. Peng, M. O. Ward, and E. A. Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In *INFOVIS'04*, pages 89–96. IEEE Computer Society, 2004.

[77] M. E. Raven and A. Flanders. Using contextual inquiry to learn about your audiences. *SIGDOC Asterisk J. Comput. Doc.*, 20(1):1–13, Feb. 1996.

[78] J. Roberts. State of the art: Coordinated multiple views in exploratory visualization. In *CMV '07*, pages 61–71, 2007.

[79] J. Rugis and P. Leary. Towards crustal reservoir flow structure modeling through interactive 3d visualization of meq & mt field data. In *In Proc. 33rd New Zealand Geothermal Workshop*, November 2011.

[80] R. I. Rusby. The future of visualization: Vision 2020. *WorldOil*, 229(1), 2008.

[81] C. Schmandt. Spatial input/display correspondence in a stereoscopic computer graphic work station. *SIGGRAPH Comput. Graph.*, 17(3):253–261, July 1983.

[82] M. Sedlmair, A. Frank, T. Munzner, and A. Butz. RelEx: Visualization for actively changing overlay network specifications. *IEEE Transactions on Visualization and Computer Graphics*, 18:2729–2738, 2012.

[83] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis)*, 18(12):2431–2440, 2012.

[84] J. Shemeta and P. Anderson. It's a matter of size: Magnitude and moment estimates for microseismic data. *The Leading Edge*, 29(3):296–302, 2010.

[85] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343, 1996.

[86] H. Siirtola and K.-J. Räihä. Discussion: Interacting with parallel coordinates. *Interact. Comp.*, 18(6):1278–1309, 2006.

[87] C. Steed, J. Swan, T. Jankun-Kelly, and P. Fitzpatrick. Guided analysis of hurricane trends using statistical processes integrated with interactive parallel coordinates. In *VAST' 2009*, pages 19–26, 2009.

[88] N. B. Sultanum. Exploring novel interfaces for 3d visualization of reservoir simulation post-processing data, August 2011.

[89] T. Taerum, M. C. Sousa, F. Samavati, S. Chan, and J. R. Mitchell. Real–time super resolution contextual close–up of clinical volumetric data. In *Proceedings of the Eighth Joint Eurographics / IEEE VGTC conference on Visualization*, EUROVIS'06, pages 347–354, 2006.

[90] Z. Ulrich. Calculating stimulated reservoir volume (srv) with consideration of uncertainties in microseismic-event locations. In *Canadian Unconventional Resources Conference*. SPE International, 2011.

[91] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[92] D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, UIST '04, pages 137–146, New York, NY, USA, 2004. ACM.

[93] A. Voida and S. Greenberg. Wii all play: the console game as a computational meeting place. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1559–1568, New York, NY, USA, 2009. ACM.

[94] D. W. Eaton, J. Akram, A. St-Onge, and F. Forouhideh. Determining microseismic event locations by semblance-weighted stacking. In *CSPG CSEG CWLS Convention*, 2011.

[95] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *AVI '00*, pages 110–119, New York, NY, USA, 2000.

[96] R. Want, A. Hopper, V. Falcão, and J. Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, Jan. 1992.

[97] M. Ward. Xmdvtool: integrating multiple methods for visualizing multivariate data. In *Visualization '94*, pages 326–333, oct 1994.

[98] W. Watney, W. Guy, J. Doveton, S. Bhattacharya, P. Gerlach, G. Bohling, and T. Carr. Petrofacies analysis-a petrophysical tool for geologic/engineering reservoir characterization. *MEMOIRS-AMERICAN ASSOCIATION OF PETROLEUM GEOLOGISTS*, pages 73–92, 1999.

[99] M. Williams and T. Munzner. Steerable, progressive multidimensional scaling. In *INFOVIS '04*, pages 57–64, Washington, DC, USA, 2004. IEEE Computer Society.

[100] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques*, pages 3–33, Washington, DC, USA, 1997. IEEE Computer Society.

[101] J. Yang, W. Peng, M. O. Ward, and E. A. Rundensteiner. Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In *Proceedings of the Ninth annual IEEE conference on Information visualization*, INFOVIS'03, pages 105–112, Washington, DC, USA, 2003. IEEE Computer Society.

[102] X. Yuan, P. Guo, H. Xiao, H. Zhou, and H. Qu. Scattering points in parallel coordinates. *Visual. and Comp. Graphics, IEEE Transactions on*, 15(6):1001–1008, 2009.

[103] N. Zamel, J. Pita, and A. Dogru. Next-generation visualization technologies for exploration & production. pages 29–31. SPE, 2001.

[104] Z. Zhang, K. T. McDonnell, and K. Mueller. A network-based interface for the exploration of high-dimensional data spaces. In *PacificVis' 2012*, pages 17–24, 2012.