CREATING OVERVIEW VISUALIZATIONS FOR DATA UNDERSTANDING

By

LI LIU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Deborah Silver

And approved by

____

____

____

____

____

____

New Brunswick, New Jersey

January, 2019

ABSTRACT OF THE DISSERTATION

Creating Overview Visualizations for Data Understanding

by Li Liu

Dissertation Director:

Prof. Deborah Silver

One of the major challenges in data visualization is the presentation problem: having too much information to display at a time in one screen. The traditional presentation techniques (e.g., panning, scrolling, and flipping) that are widely used in standard user interfaces always introduce a discontinuity between the information displayed at different times and places. Viewers find a compact visual summarization of the information space (i.e., an overview) is helpful in data understanding. When utilized properly, an overview can provide users with an immediate appreciation and an overall sense of data. Although creating an overview is often a design goal and an overview is widely noted in data visualization as a qualitative awareness of one aspect of the

data (e.g., gaining an overview of the information space) or a technical and user interface component (e.g., overview + detail visualization), the properties and categories of overviews, the relations between overviews and viewers' awareness, and the process to create overviews are barely discussed in the literature.

In data visualization, giving an overview of a dataset is part of a broad topic of providing a combination of contextual and detailed views. Although discussions on contextual and detailed visualizations are mostly made in the scope of information visualization, approaches are also used in scientific visualization. These visualizations are studied and classified by interface mechanisms (e.g., overview+detail, focus+context, contextual cue) used to separate and blend views without considering the characteristics of information space. More importantly, not all of the contextual and detailed visualizations give an overview of data.

In this thesis, I focus on "overview" visualizations as a means to covey the context of a large dataset. An overview visualization is a visual representation that provides viewers with an overall awareness of the content, structure, or changes of the data (the dataset can contain time-varying information) while allowing the viewer to further drill down into the details. The applicability of overview visualizations is extended to both scientific visualization and information visualization domains. Overview visualizations are characterized into five important aspects: (1) the nature of overviews; (2) the roles of overviews; (3) design strategies for the overview display; (4) approaches for shrinking the information space of data; (5) techniques for the interactions and the detail representations. Based on the characterization of overview visualizations and inspired by other visualization design models, a pipeline model is created to analyze existing systems or papers and

to guide the development process of an overview visualization. Two case studies are presented with evaluations to illustrate the general usage of the proposed characterization and pipeline model. The first involves a time-series 3D dataset of ocean simulations (scientific visualization) and the second involves a university career job portal (information visualization). The results demonstrate how overview visualizations can facilitate data understanding and analysis.

Acknowledgements

I would like to thank Prof. Deborah Silver for her vision and encouragement throughout my years in Vizlab at Rutgers University. It is her support and expert guidance that makes this thesis possible. I also want to thank Prof. Min Chen for his valuable input and suggestions to my research.

I am very grateful to Dr. Karen Bemis who got me started on my research and continuously gave me advise, encouragement, and proof-reading of my papers and thesis.

I wish to thank my collaborators: Prof. Enrique Curchitser and Dr. Dujuan Kang in the Department of Environmental Sciences at Rutgers; career counselors in the Rutgers Career Services; Dr. Carolyn Philip and Dr. Tom Peterka in the Argonne National Lab.

I wish to thank my co-students in the Vizlab – Sedat Ozer and Jay Takle. Collaborating with them on research projects has helped me gain experience to work on different types of problems.

I wish to thank all my friends in the ECE department, teachers who have taught me, and the friends I have made over the years at Rutgers. It is an experience to remember.

Finally, I want to thank my parents Yongming Liu and Jie Qin for all their support and encouragement during my studies.

Table of Contents

ABSTRACT OF THE DISSERTATION ii

Acknowledgements v

Chapter 1: Introduction 1

1.1 Presentation Problem in Data Visualization 1

1.2 Contextual and Detailed Visualizations 3

1.3 Overviews in Data Visualization 5

1.4 Limitations of Current Research 6

1.5 Summary of Contributions 7

1.6 Organization of the Thesis 8

Chapter 2: Related Work 10

2.1 Reviews of Contextual and Detailed Visualizations 10

2.2 Visualization Models 12

Chapter 3: A Characterization of Overview Visualizations 14

3.1 Dimension 1 - The Nature of Overviews 15

3.1.1 Content 15

3.1.2 Structure 17

3.1.3 Changes 18

3.2 Dimension 2 - The Roles of Overviews 19

3.3 Dimension 3 - Design Strategies for Overview Display **19**

3.3.1 Time-centric Design 20

3.3.2 Space-centric Design 21

3.3.3 Spatiotemporal-centric Design 23

3.3.4 Abstract Context Design 24

3.4 Dimension 4 – Approaches for Shrinking the Information Space of Data **25**

3.5 Dimension 5 – Techniques for Interactions and Detail Representations **26**

3.5.1 Detail-In-Overview 26

3.5.2 Detail-Out-Of-Overview 31

Chapter 4: A Pipeline Model for The Development of Overview Visualizations **33**

4.1 Analyze Domain Problem and The Related Data **33**

4.2 Identify the Basics of Overview and Design Principles + Characterize Data **35**

4.3 Design the Overview Display + Transform Data **36**

4.4 Design the Interactions and Detail Display + Compute Metadata **37**

4.5 Implement the Overview Visualization **38**

Chapter 5: Introduction to The Case Studies **39**

Chapter 6: Case Study 1 - Illustrative Visualization of Mesoscale Ocean Eddies **41**

6.1 Background **42**

6.2 Related Work **46**

6.3 Data and Approach **49**

6.3.1 ROMS Data ... 49

6.3.2 Approach Overview ... 50

6.4 Eddy Feature Computation ... 51

6.4.1 Eddy Feature Extraction ... 51

6.4.2 Eddy Path Detection ... 55

6.4.3 Properties Calculation ... 57

6.5 Illustrative Eddy Visualization ... 59

6.5.1 Skeletonized Eddy Path Map ... 60

6.5.2 Three-dimensional Eddy Path Map ... 61

6.5.3 Isolated Eddy Path Visualization ... 63

6.5.4 Individual Eddy Visualization ... 63

6.6 Expert Evaluation ... 66

6.6.1 Methodology ... 67

6.6.2 Expert Feedback ... 67

6.7 Conclusion ... 70

Chapter 7: Case Study 2 – Interactive Visualization of Job Posting Data ... 72

7.1 Background ... 73

7.2 Related Work ... 75

7.3 System Design ... 77

7.3.1 Data Characterization ... 78

7.3.2 Tasks Abstraction ... 79

7.3.3 Design Principles and Visual Encoding ... 80

7.3.4 User Interface and Interactions ... 83

7.3.5 System Overview ... 85

7.4 Algorithm Design ... 86

7.4.1 Preprocessing ... 86

7.4.2 Visualization Layout Creation ... 87

7.4.3 Major-Career Links Selection ... 88

7.4.4 Tag Cloud Computation ... 89

7.5 User Scenarios ... 89

7.5.1 Scenario One ... 89

7.5.2 Scenario Two ... 91

7.6 User Study ... 92

7.6.1 Questionnaire Details ... 93

7.6.2 Study Protocol ... 94

7.6.3 Results ... 96

7.6.4 Discussion of User Study ... 98

7.7 Lessons Learned ... 99

7.8 Conclusion ... 102

Chapter 8: Discussions ... 103

8.1 General Usages of the Pipeline Model in Case Studies 103

8.1.1 Case Study One 103

8.1.2 Case Study Two 107

8.2 Observations from Case Studies 111

Chapter 9: Conclusion and Future Work 118

Chapter 10: Summary of Accomplishment and Related Publications 119

Reference 121

Appendix A: Implementation of Case Study 1 139

A.1: Computation Performance 139

A.2: Eddy Computation Program 140

A.3: Formats of the Output Files 140

A.4: Accesses to the WebGL Eddy Visual Explorer and the Data for Visualization 144

Appendix B: Implementation of Case Study 2 145

B.1: Scalability Analysis 145

B.2: Source Code 146

B.3 License 147

Appendix C: Accompanying Videos 147

Chapter 1: Introduction

1.1 Presentation Problem in Data Visualization

Data visualization involves the creation and study of the visual representation of data. There are numerous situations in which we use data visualization to understand data. This could involve anything from last week's stock market trends to traffic accident reports or even weather forecasts for various geographical areas. Thanks to the great bandwidth of the human visual system, data visualization provides an efficient communication tool in illustrating concepts, ideas, and properties intrinsic to the data. Data can have a wide variety of forms, but typically follow one of two types: (1) data that has a natural geometric structure and correspondence in physical space; and (2) abstract data that doesn't have an intrinsic geometric structure. In visually representing the data of first type, we speak of scientific visualization. Scientific visualization is a discipline that aims to visually represent the results of scientific experiments or natural phenomena. Abstract data often result from some activity generated by humans, but do not correspond to a physical object positioned in any part of space. The visual representation of abstract data is referred to as information visualization.

One of the challenges in data visualization is how to allow an end-user to work with, navigate through, and generally analyze a set of data that is too large to fit in a display. The problem may be caused by displaying too many cases and variables in a screen (technological factor) or viewers may only be able to highlight particular cases or particular variables (human factor) at a time.

Common ways of browsing large datasets include zooming, scrolling, panning, and flipping, which are generally default interaction mechanisms in most computers or other devices with screens. These browsing methods only present a portion of the data at a time and thus introduce a discontinuity between the information displayed at different times and places. Figure 1.1 gives three examples of browsing different datasets with zooming, scrolling, and flipping.

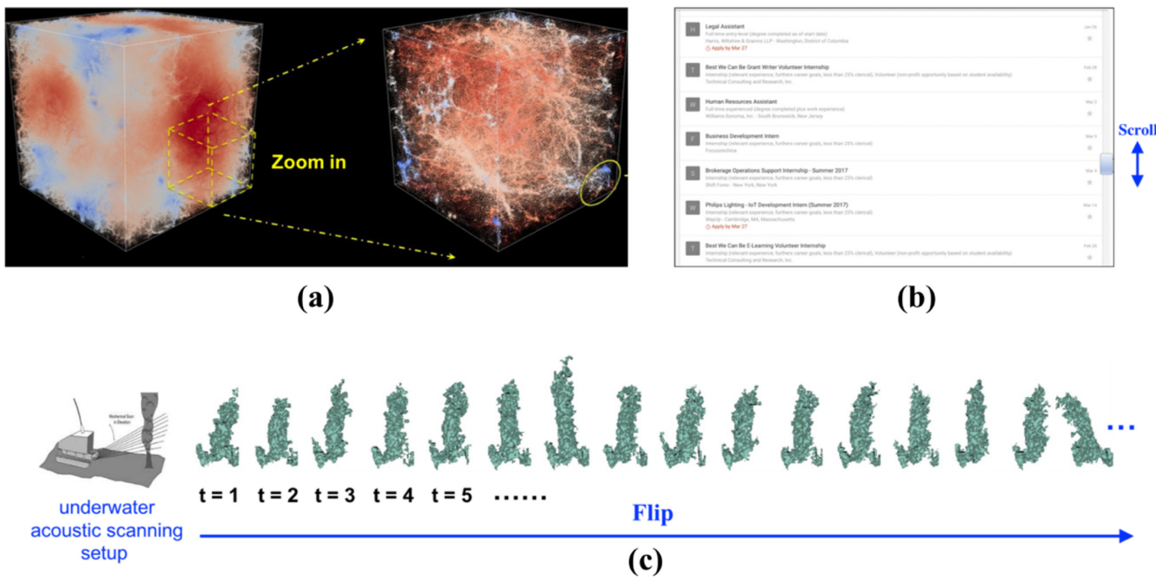


Figure 1.1: Examples of browsing large datasets with traditional ways. (a) Zoom in the display of universal black matter halo [THP12] to get more detail of a certain region; (b) Scroll down a list of job postings; (c)Flip through a time series of acoustic hydrothermal plume on the seabed [BSX15].

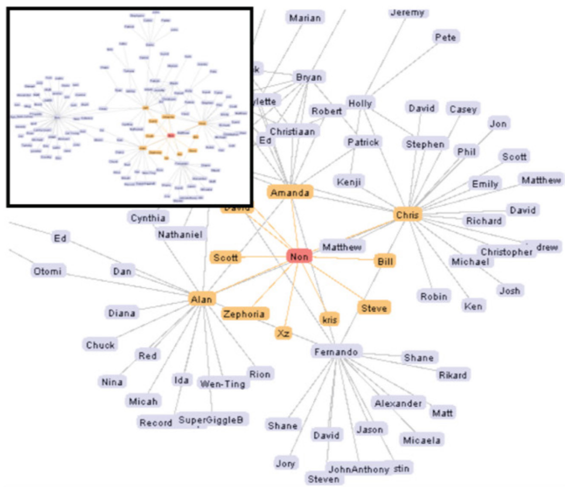
Another way to present and understand large quantities of data is to give an overview of the data. The famous information-seeking mantra from Shneiderman et al. [Shn96] reads: "Overview first, zoom and filter, details on demand." Showing an overview of a dataset can help viewers stay oriented while drilling down into the details. This becomes a design goal in many visualizations. Considerable effort has been devoted to creating overviews and developing useful techniques.

However, these efforts focus on specific application domains rather than a systematic study on overview visualizations. This thesis is motivated by the following questions: What is the definition of overview visualizations? What are the examples of overview visualizations? How do we categorize and characterize overview visualizations? What is the process of creating overview visualizations?

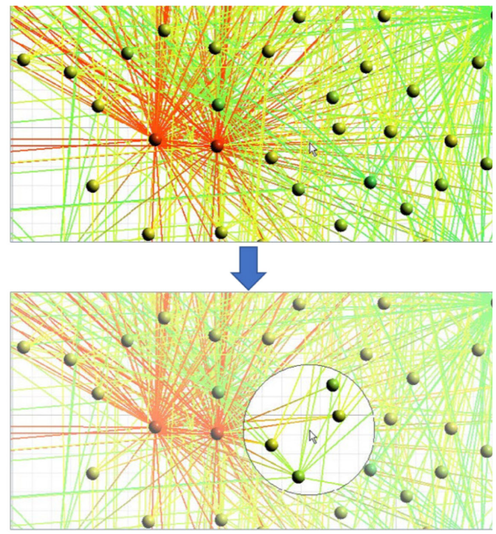
1.2 Contextual and Detailed Visualizations

In data visualization, giving an overview of a dataset is part of a broad topic of providing a combination of contextual and detailed views. This thesis refers to these techniques as contextual and detailed visualizations. Contextual and detailed visualizations offer potential advantages derived from their ability to allow users to rapidly and fluidly move between contextual views and detailed views [CKB08]. On one hand, a contextual view can help present overall patterns, assist user with navigation and search, and orient activities. On the other hand, a detailed view allows users to examine details such as individual cases or variables. According to the interaction mechanisms used to separate and blend views, contextual and detailed visualizations can be broadly categorized into four types: a. overview+detail, which spatially separates focused and contextual information in different views at varying levels of detail; b. zooming, which involves changing the scale of the viewed area by temporal separation; c. focus+context (also known as detail-in-context), which integrates focus (detailed information) and context (contextual information) into a single display by using spatial distortions, lenses, or over lays to efficiently use the available screen space;

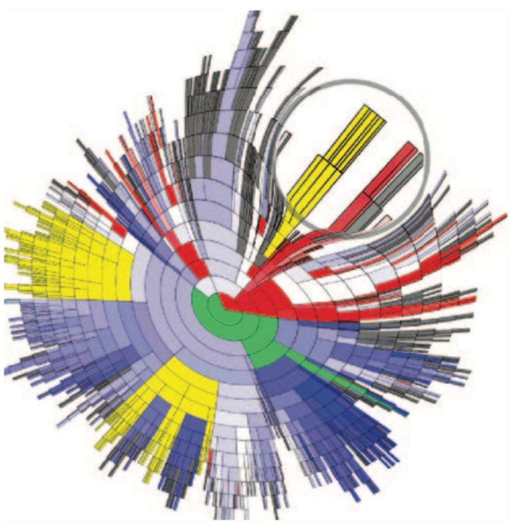
and d. contextual cue which uses supplemental cues that highlight salient items within the information space. Figure 1.2 illustrates each these types of visualizations with an example.



(a)



(b)



(c)

(d)

Figure 1.2: Four schemes of contextual and detail visualizations. (a) Overview+Detail (Prefuse [HCL05]); (b) Zooming (Graph Lens [TAHS06]); (c) Focus+Context (Radial Clustergram [ABF07]); (d) Contextual Cues (SDOF [KMH02]).

1.3 Overviews in Data Visualization

In data visualization research and practice, overview is a frequently used notion and design goal. At least two uses of the term overview are found in the literature. The first use is about viewers gaining an overview of the information space. As Spence noted in his book [Spe14], the term overview "implies a qualitative awareness of one aspect of some data, preferably acquired rapidly and, even better, pre-attentively: that is, without cognitive effort". Metaphoric phrases like "bird's eye view", "gestalt view", "broad overview", "global overview", and "big picture" also appear in literatures to indicate a similar understanding [HH11]. The term "overview" is also used in the sense of a technical and user interface component as in the phrase "overview+detail visualization". In this thesis, unless used in the combination "overview+detail", an overview is defined as a compact visual representation of the information space which summarizes some commonalities of attributes in the data so as to assist users in navigating the information space and orienting patterns and activities in the data.

An effective overview can provide users with an immediate appreciation and an overall sense of the content, structure, and dynamics in the data. Well-designed overviews will also provide users with directly manipulable control mechanisms that facilitate exploration and allow users to focus attention on the larger information problem at hand [GMPS00]. For example, in Figure 1.3 a full-text visualization of the Iraq War Logs gives an overview of 11, 616 WikiLeaks reports from 2006 [Str10]. Representing reports as dots labelled with "characteristic" words and linking "similar" documents with edges draws an overview of the Iraq War as described in these reports.

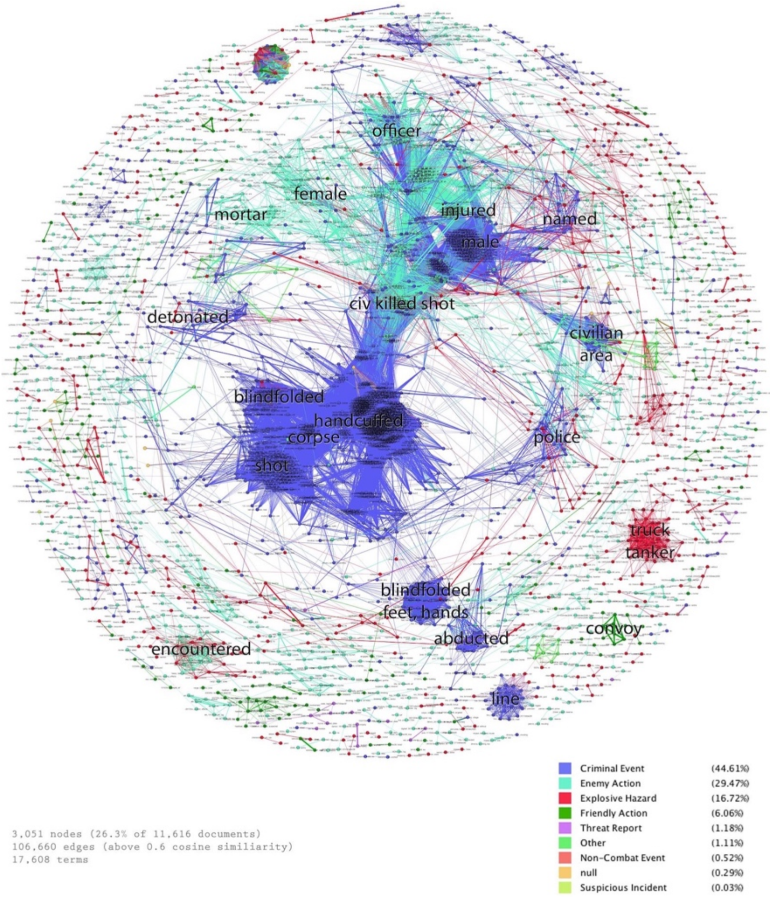


Figure 1.3: A full-text visualization of the Iraq War logs constitutes an example of the overview in data visualization [Str10]. This visualization is created from 11,616 SIGACT WikiLeaks reports from December 2006, the bloodiest month of the war. Each report is a dot that is labelled by the three most "characteristic" words in that report. Documents that are "similar" have edges drawn between them. The location of the dot is abstract, and has nothing to do with geography. Instead, dots with edges between them are pulled closer together. This produces a series of clusters, which are labelled by the words that are most "characteristic" of the reports in that cluster.

1.4 Limitations of Current Research

Current studies on the contextual and detailed visualizations cannot answer the questions of defining and categorizing overview visualizations: (1) the discussions are mostly made in the scope of information visualization but the techniques are also used in scientific visualizations (e.g., interactive lenses [TGK17]); (2) the categorizations of contextual and detailed visualizations are according to the interface mechanisms that are used to separate and blend views (e.g. overview + detail, zooming, etc.) and barely consider the construction of information space and the nature of data; (3) not all of the contextual and detailed visualizations give an overview of the data.

Although overview is made a design goal and is widely noted in the literatures of information visualization, the properties and categories of overviews, the relations between overviews and viewers' awareness, and the process to create overviews are barely discussed in the literatures.

Furthermore, many visualization models have been proposed to guide the creation and analysis of visualization systems [CM97, CR98, PAK08, Mun09], but they are not tightly coupled to the question of creating overview visualizations.

1.5 Summary of Contributions

This thesis addresses the above noted limitations in current research by making the following contributions:

(1) Define overview visualization and extend its applicability to cover scientific visualization as well as information visualization.

(2) Characterize overview visualizations into five important aspects and propose a pipeline model for analyzing or developing overview visualizations.

(3) Present two case studies (with evaluations) of creating overview visualizations for data understanding in scientific visualization and information visualization.

(4) Illustrate the general usage of the proposed characterization and pipeline model in the case studies and demonstrate how overview visualizations can help data understanding and analysis.

1.6 Organization of the Thesis

This thesis is organized as follows:

Chapter 2 reviews the related work including previous reviews on the contextual and detailed visualizations and other visualization models.

Chapter 3 gives a definition of overview visualization, characterizes overview visualization into five important aspects, and classifies and elaborates each of the aspects.

Chapter 4 proposes a pipeline model for the creation of overview visualizations based on the characterization of overview visualization described in Chapter 3.

Chapter 5 gives a brief introduction to the two case studies used in the thesis.

Chapter 6 is the first case study: illustrative visualization of ocean eddies. The content of this case study was published in the journal of Computer Graphics and Forum (paper title "Illustrative Visualization of Mesoscale Ocean Eddies" [LSB17]).

Chapter 7 is the second case study: visualization of job posting data. The content of this case study was published in the journal of IEEE Computer Graphics and Application (paper title "Application Driven-Design: Help Students Understand Employment and See The "Big Picture" [LSB18]).

Chapter 8 first discusses the general usages of the proposed characterization and pipeline model of overview visualizations in the case studies, then derives some observations about overview visualizations from the case studies.

Chapter 9 concludes the thesis and presents future work.

Chapter 10 lists the accomplishments and related publications of the author.

Chapter 2: Related Work

2.1 Reviews of Contextual and Detailed Visualizations

In data visualization, overview visualizations of data are discussed in a broad topic of contextual and detailed visualizations. Previous reviews often give categorizations of contextual and detailed visualization according to the interface mechanisms that are used to separate and blend views. Leung and Apperley [LA94] provided the first comprehensive survey and taxonomy on distortion-oriented presentation techniques. The authors classified the techniques in terms of their magnification functions. Shneiderman et al. [Shn96] proposed a taxonomy of information visualization based on seven data types and seven tasks that information visualizations may support. He focuses on design advice for designers of visualizations and on tasks information visualizations may support, especially the tasks of gaining an overview. Hauser [Hau06] proposed a generalized definition of focus+context visualization extending its applicability to scientific visualization. The author showed how different graphics resources such as space, opacity, color, etc., can be used to visually discriminate between data subsets in focus and their respective context. He also discussed the interaction aspect of focus+context visualization. Cockburn et al. [CKB08] gave a thorough review on the existing contextual and detailed visualization techniques and categorized these techniques into overview+detail, zooming, focus+context, and technique-based techniques according to the inferface mechanisms used to separate and blend views. Critical features of these categories, and empirical evidence of their success, are discussed in the paper. The categorization is widely adopted and cited in different publications.

Reviews have also addressed different application domains. Plaisant et al. [PCS95] described user tasks for image browsing, considering alternative interfaces for overview+detail, zooming, and focus+context approaches. Herman et al. [HMM00] reviewed graph visualization and navigation schemes, including focus+context views, with an emphasis on graph layout algorithms. Robert [Rob07] provided a "state of the art" of multiple coordinated views that enables users to visualize their data. The review describes areas that should be developed further and looks at what the future may hold for multiple coordinated views in exploratory visualization. Burigat et al. [BC13] gave a detailed analysis of the literature on overview+detail visualization and compares the results of desktop and mobile devices studies to highlight strengths and weaknesses of the approach. Then, the paper presents an experiment that studies unexplored aspects of the design space for mobile interfaces based on the overview+detail approach, investigating the effect of letting users manipulate the overview to navigate maps and the effect of highlighting possible objects of interest in the overview to support search tasks. Radeva et al. [RLH14] proposed a generalized temporal focus+context framework to classify already existing common techniques and to define new techniques that can be used in medical volume visualization. The new techniques explore the time-dependent position of the framework focus region to combine 2D and 3D rendering inside the focus and to provide a new focus-driven context region that gives explicit spatial perception cues between the current and past regions of interest. Tominski et al. [TGK14] [TGK17] surveyed the literature on interactive lenses in the context of visualization.

In the characterization of overview visualization proposed in this thesis (Chapter 3), the fifth dimension, which is the techniques for interactions and detail representations, adopts some of the categorizations from these reviews of contextual and detailed visualizations.

2.2 Visualization Models

Creating overview visualizations can be treated as a problem of visualization design. Visualization models have been proposed to guide the creation and analysis of visualization systems as in [CM97, CR98, PAK08, Mun09]. In [CM97], Card et al. sketched a scheme for mapping the morphology of the design space of visualizations, which resulted in a framework for designing new visualizations and augmenting existing designs. Chi et al. [CR98] proposed a framework to include operators and interactions in visualization systems. The framework enables a new way of exploring and evaluating the design space of visualization operators and helps end–users in their analysis tasks [CR98]. Purchase et al. [PAK08] presented three different approaches to theoretical foundations of Information Visualization: data-centric predictive theory, information theory, and scientific modeling. An over-arching framework for these three approaches is provided based on the definitions from linguistic theory [PAK08]. Munzner [Mun09] presented a nested model for the visualization design and validation with four layers. This nested model provides prescriptive guidance for determining appropriate evaluation approaches by identifying threats to validity unique to each level [Mun09].

Although these visualization models are not tightly coupled to the question of creating overview visualizations, they heavily influenced the pipeline model proposed in this thesis (Chapter

4). Firstly, the pipelines in the proposed model are inspired by the visualization pipeline described in [CM97] and the visualization operator model in [CR98]. Second, the relation between visualization design and data flow in the proposed model is similar to that in [PAK08]. Finally, the connections between different levels in the proposed pipeline model are similar to those of the nested model [Mun09].

Chapter 3: A Characterization of Overview Visualizations

In this thesis, an overview visualization is defined as: an interactive visualization that shrinks an information space of data to a coarse level of granularity to provide visual summarizations of content, structure, or dynamics of the data while keeping the capability of showing more details through user interactions. Rather than proposing a single taxonomy of overview visualization, this thesis characterizes overview visualizations into five dimensions: (1) the nature of overviews; (2) the role of overviews; (3) design strategies for the overview display; (4) approaches for shrinking the information space of data; and (5) techniques for the interaction and the detail representation. Corresponding to the important aspects of overview visualizations, these characterized dimensions guide the design and analysis of overview visualizations from top to bottom by helping viewers to determine: What it means to have an overview? What roles does an overview play in the visualization? How an overview is constructed from data? How overviews and details are displayed and controlled through interactions. The characterization is a unified taxonomy model where each dimension has its own taxonomies. Figure 3.1 sketches the proposed characterization of overview visualizations. The rest of this chapter will elaborate all of the characterized dimensions and their taxonomies with examples.

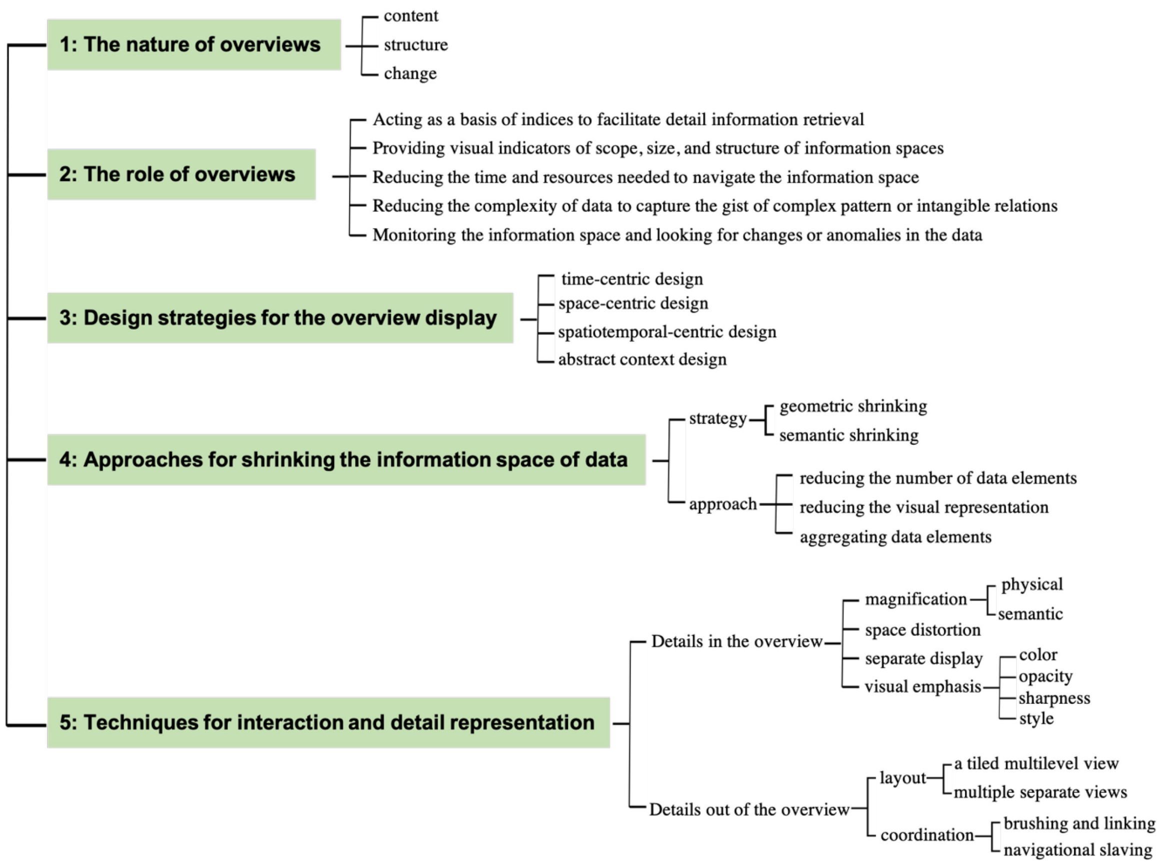


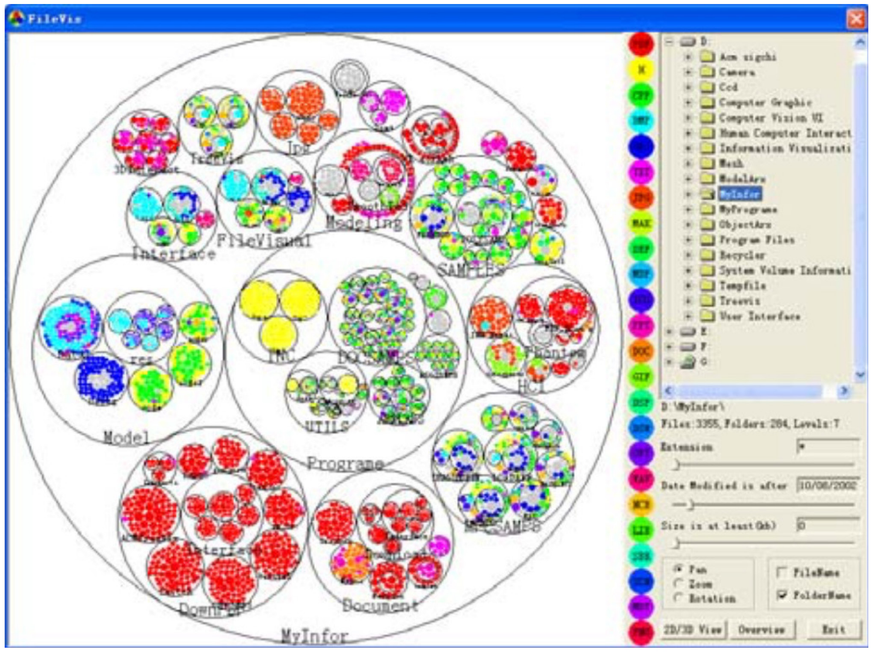
Figure 3.1: The characterization of overview visualizations.

3.1 Dimension 1 - The Nature of Overviews

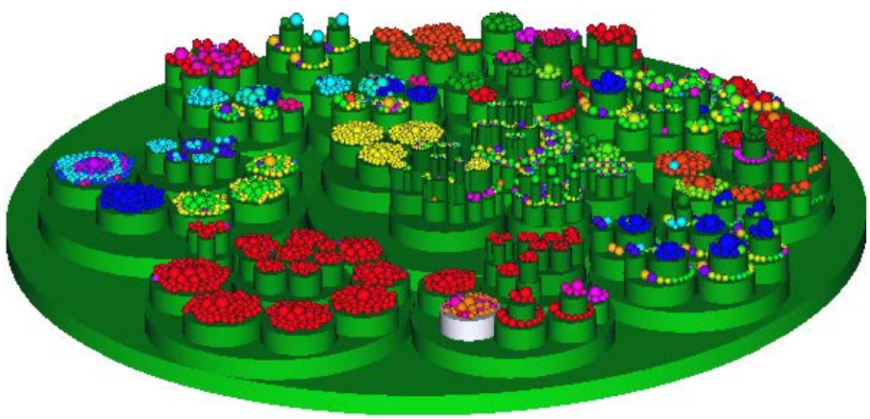
The first dimension specifies the nature of the information that one is aware of when having an overview. Based on the awareness that an overview gives to a viewer, overview visualizations can be categorized into three types: content, relation, and change.

3.1.1 Content

In this type, overviews are the subject matter of the information space. Viewers' awareness from this type of overviews is mainly about the content of a large set of data, a subset of the information space, or the information space itself. Figure 3.2 shows a visualization that gives an overview of a



(a)



(b)

Figure 2.2: An example of overview visualization that gives an awareness of the content of data [WWDW06]. (a) The user interface of a file system visualization shows an overview of the directory "D:\MyInfor" using nested circles. Directories are represented by white circles in which files are symbolized as colorful circles. The radii of colorful circles indicate the file sizes. (b) 3D nested cylinders and spheres of (a).

file system [WWDW06]. The content in the file system (i.e., directories and files) are all represented by nested circles with different colors.

3.1.2 Structure

Overviews of this type are about patterns, relations, and structures among the data, for instance among variables or within a set of information objects. Sometimes, overviews concern the context of a data set or more likely the context of a part of the data. In [GWG03], one visualization is characterized as providing a good overview of the structure of a shared workspace. Another example shown in Figure 3.3 is an overview of a biclustering with 20 clusters [SGG14].

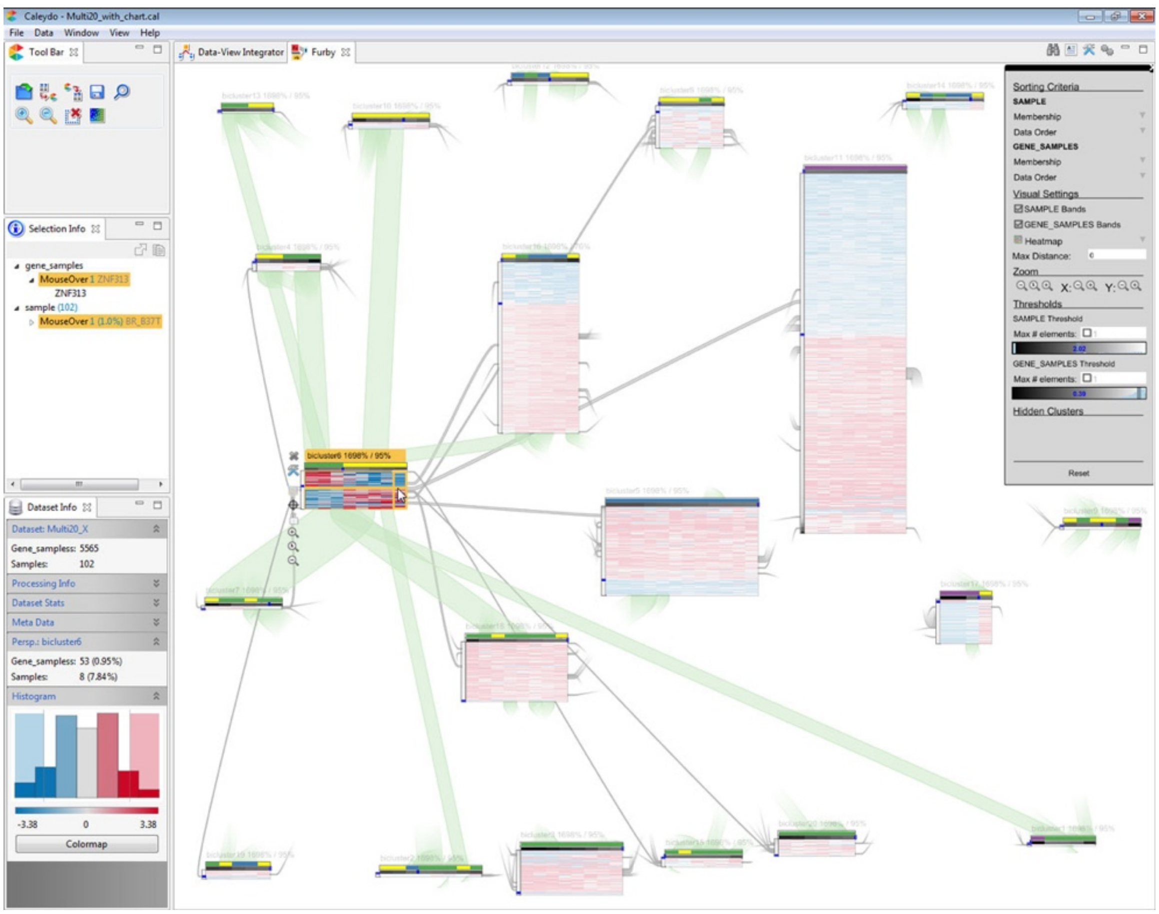
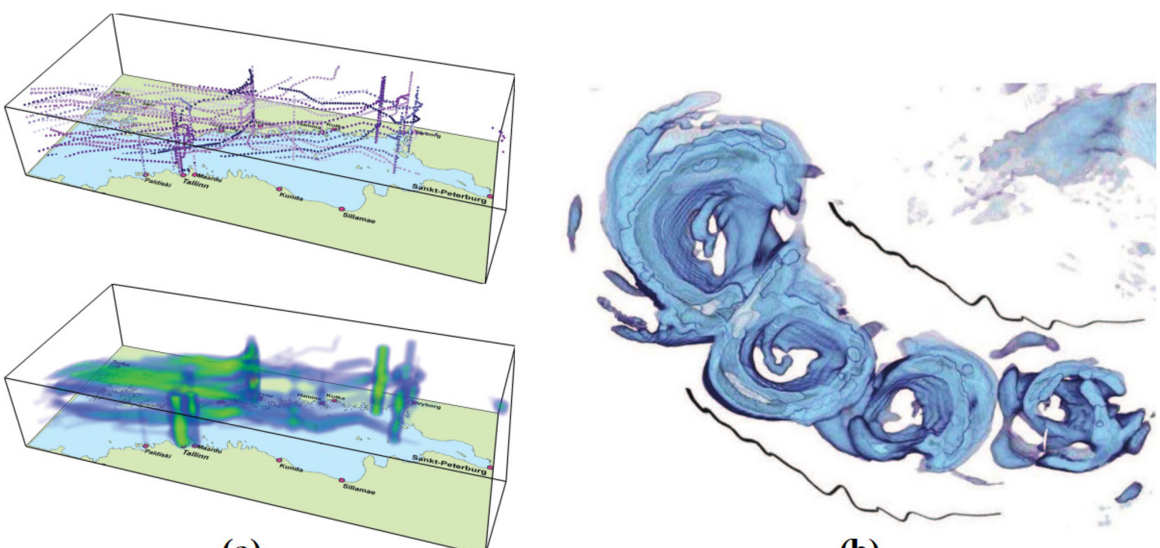


Figure 3.3: An example of overview visualization that gives an awareness of the structure or patterns in data [SGG14]. This figure shows an overview visualization of a biclustering result with 20 clusters. Grey bands show the overlap in gene dimension and green bands visualize the relationships in sample dimension.

3.1.3 Changes

This type of overview provides viewers with an awareness of the shifting of normal state or changes of content and/or structure in the information space. An example is presented in Figure 3.4a which summarized the tracks of multiple moving objects (ocean vessels) by trajectories (the top of the figure) or volume rendering techniques (the bottom of the figure) [AAD10]. Another example is visualizing hurricanes using illustration-inspired techniques [JCRS09] as shown in Figure 3.4b. The visualization gives an overview of the intensification process of Hurricane Isabel and uses speed lines and silhouette to convey the motion of the hurricane.



(a) **(b)**

Figure 3.4: Examples of overview visualizations that provide viewers with awareness of changes in the data. (a) Aggregation of trajectories summarize the movements of tankers. Top: trajectories are depicted as traces in the space–time cube. Bottom: the tanker movement is shown by using volume rendering technique [AAD10]. (b) This illustrative visualization gives an overview of the intensification process of Hurricane Isabel. Silhouettes and speed lines are used to convey the motion of the hurricane [JCRS09].

3.2 Dimension 2 - The Roles of Overviews

The second dimension is the role of overviews in the visualization, which is about the purposes of creating an overview or the tasks that the created overview supports. An overview can play one or more roles in the visualization. The main roles for overviews include:

R1. Aiding details retrieval. In this case, the overview is used as a basis of indices (or references) where details can be easily identified and accessed.

R2. Providing visual indicators of scope, size, structure, or other properties of data to reduce the rendering cost of visualization tasks and the visual burden of viewers. The overview acts as a surrogate for a large amount of data elements.

R3. Reducing the time and resources needed to navigate the information space of data by showing a limited number of selected objects. The overview serves as a preview of the data.

R4. Reducing the complexity of data to support viewers' need to capture the gist of complex heterogeneous patterns or intangible relations in the data.

R5. Monitoring the information space and looking for changes or anomalies in the data.

3.3 Dimension 3 - Design Strategies for Overview Display

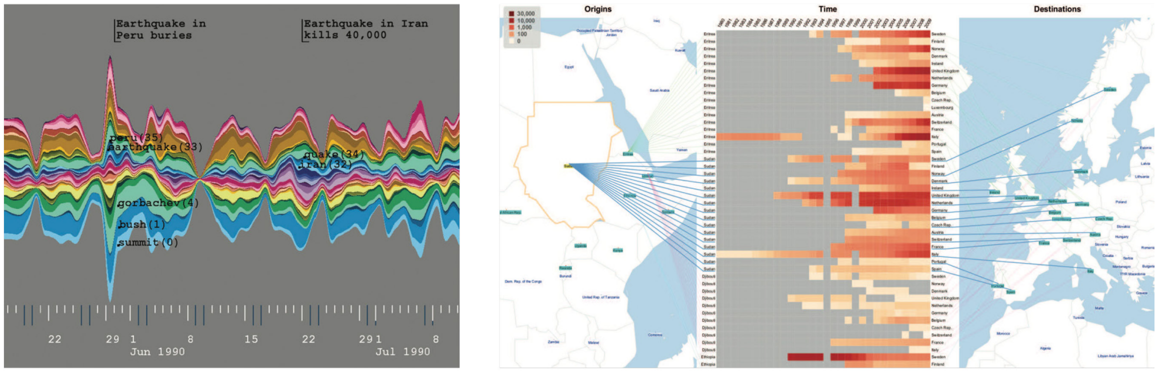
This dimension considers design strategies for transforming the information space of data into an overview display. The design choices are mostly domain and data-dependent. Design strategies for overview display generally fall into one of the following categories: time-centric design, spatial-centric design, spatiotemporal-centric design, and abstract context design [AWB17].

3.3.1 Time-centric Design

This design strategy maps temporal information to one of the spatial dimensions of the display and focuses on the variation of attribute values in the temporal context. It typically employs a conventional time series visualization (e.g. line plots and calendar view) as the base representation augmented with other information (e.g., spatial information) in abstract forms or available on demand [AWB17]. For spatial data, this approach suppresses spatial information to a reduced style or non-spatial visual channel, which usually incurs a loss of spatial information and often introduces difficulties in spatial awareness and reasoning. Since the absence of intrinsic spatial mapping of spatial data can be problematic, time-centric design mostly involves mapping a non-spatial data to the temporal dimension.

An example is ThemeRiver [HHN00] which uses a 'river' metaphor to show changes in multivariate data over time. Figure 3.5a shows a ThemeRiver representation of media mentions of news concerning Cuba around 1960 and 1961. Each topic is encoded as a colored 'current' whose width can represent the number of reported items. The river itself provides an overview of the multivariate data, but the annotation shows details of the content of each current as well as significant events occurring during the selected time period [HHN00].

Another example is Flowstrates [BBB11] which uses a heatmap to provide an overview of the temporal origin-destination data and allows focusing on a region or a specific location for a specific time period. Figure 3.5b shows flows of refugees between East Africa and Western Europe. The heatmap shows the flow magnitudes by year and origin-destination. By following the lines of the



(a) **(b)**

Figure 3.5: Examples of overview visualizations with time-centric design strategy. (a) A ThemeRiver representation of media mentions of news concerning Cuba around 1960 and 1961. The river provides an overview of the changes in the data over time [HHN00]. (b) A Flowstrates representation constitutes an overview visualization of flows of refugees between East Africa and Western Europe and the changes of the magnitudes over time [BBB11].

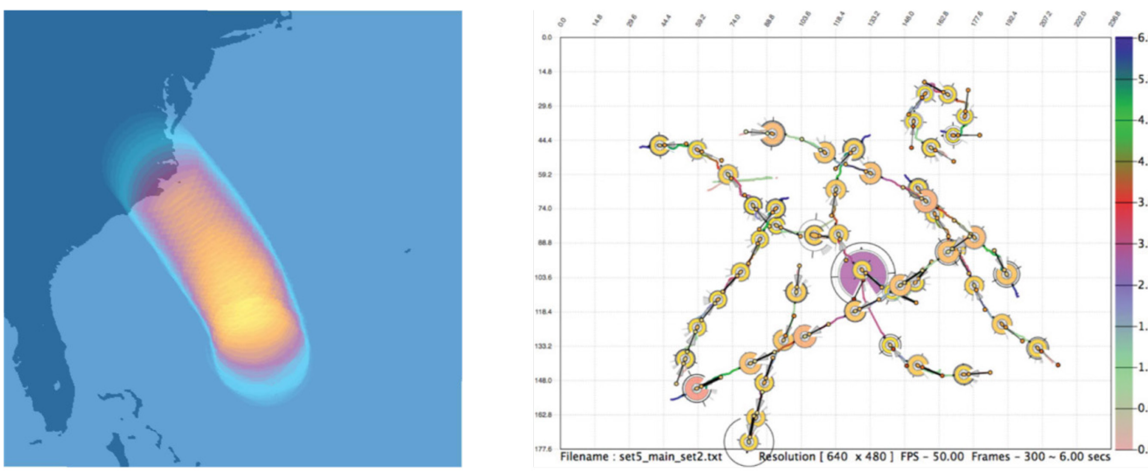
heatmap it is possible to see the flows' origins, destinations and the changes of the magnitudes over time. Users can perform spatial visual queries, focus on different regions of interest for the origins and destinations, and analyze the changes over time.

3.3.2 Space-centric Design

This design strategy presents the variation of attribute values in a spatial context and gives a principal attention to the spatial dimensions. Typically, time is depicted using animation rather than any spatial bandwidth of the display. Many visual designs in this family use a spatial visualization (e.g. a map) as the background and overlay temporal information (e.g. tracking information) on top of the background. For non-spatial data, a spatialization process is usually applied. Spatialization is a generic methodology in visualization for mapping non-spatial attributes (e.g. weight, accuracy, and risk) to spatial variables (e.g. length, size, and location) [AWB17].

An example of space-centric design is the multitemporal visualization of time-varying adaptive mesh refinement data [GABJ08]. The visualization constructs an overview from synchronizing all timesteps or a select subset of timesteps through the aid of a generated composite template. Colors are used to convey the important features from all synchronized timesteps. In Figure 3.6a, the overview visualization depicts summary statistic information gathered from queries processed over 48 timesteps from the Hurricane Isabel dataset. In the visualization, yellow regions indicate where low pressure predominantly exists across the 48 timesteps [GABJ08].

Another example is the glyph-based video visualization for semen analysis [DTW15]. This approach provides domain scientists with a summarization of the motions in the video in order to observe and compare some a large collection of spatiotemporal measurements. As shown in Figure



(a) **(b)**

Figure 3.6: Examples of overview visualizations with space-centric design strategy. (a) An overview visualization depicts summary statistic information gathered from queries processed over 48 timesteps from the Hurricane Isabel dataset [GABJ08]. (b) An overview visualization provides a summarization of the motions in the video. The motion characteristics of a set of cells along their tracks are represented as glyphs in a spatial context [DTW15].

3.6b, the motion characteristics of a set of cells along their tracks are represented as glyphs in a spatial context.

3.3.3 Spatiotemporal-centric Design

This design strategy depicts both spatial and temporal contexts on a display and attempts to give a similar level of attention to both space and time with some abstraction, omission, or distortion. Many visualization designs of this family adopt a perceptual distortion approach in 3D-to-2D projection [AWB17].

An example is video visualization which summarizes video sequences using volume visualization techniques to help users obtain an overview of a video. Figure 3.7a shows an overview visualization of relative and absolute difference volumes of a car park video by extracting interesting information from large volumes of video data [DC03]. The visual representations of the difference indicate the level of activities during the recording period, that is, movement of cars.

Another example is the multifield video visualization based on the actions of objects [BBS08]. The visualization is design to make it easy for viewers to gain an overview of a temporal segment of a video without watching the video or trying to piece together an overview from several disconnected snapshots. Figure 3.7b is a visualization of all activities of the object in a surveillance video. The display of extracted video attributes is combined with original video frames to construct an overview. A thin curve is used to indicate the center of the object's position.

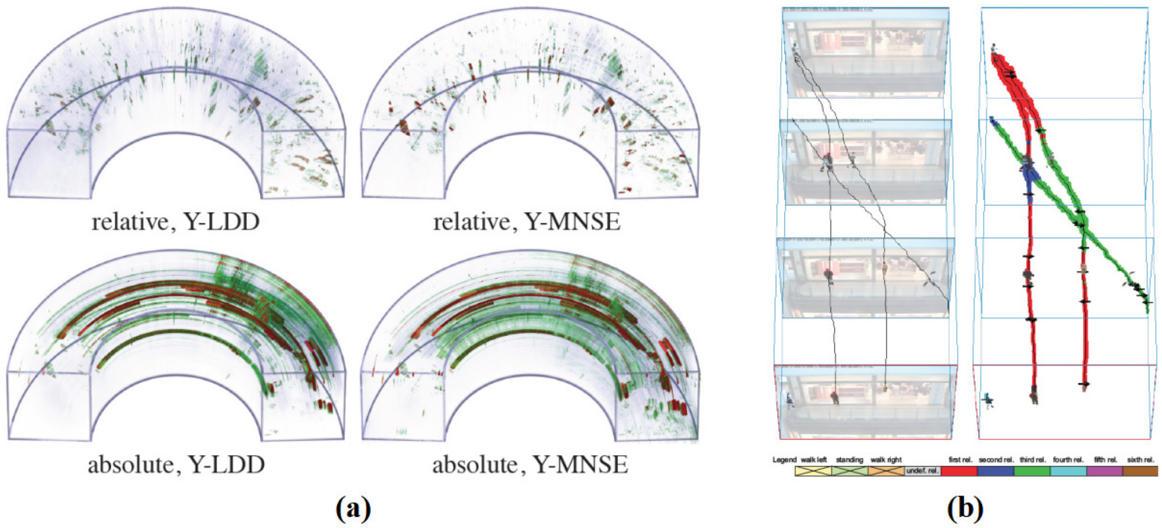
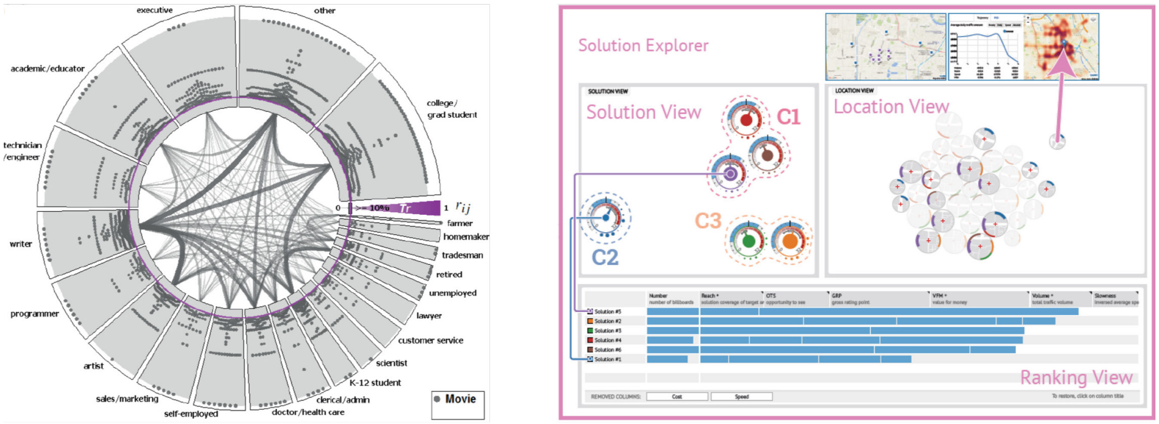


Figure 3.7: Examples of overview visualizations with spatiotemporal-centric design strategy. (a) An overview visualization of relative and absolute difference volumes of the car park video [DC03]. (b) An overview visualization of all activities of the object in a surveillance video [BBS08].

3.3.4 Abstract Context Design

This design strategy is mostly used in information visualization and is applied to non-spatial (abstract) data. An abstract context is a data graph that is constructed by using abstract representations of a set of coordinates and the associated data values. A variety of visualization applications have been developed to provide an overview of data in an abstract context. In the Contingency Wheel [AAMG12], individual items (e.g., movies) are aggregated into a connected circular layout to give a summarization view of an asymmetrically 2-way contingency table (Figure 3.8a). Another example is SmartAdP [LWL17] as shown in Figure 3.8b. The visualizations lay out problem solutions as glyphs in an abstract context (i.e., solution view, location view, and ranking view) to give overviews of a huge solution space and the relationships among these solutions to help users compare the solutions in a visual and intuitive manner.



(a) **(b)**

Figure 3.8: Two examples of overview visualizations with abstract context design. (a) Contingency Wheel [AAMG12]; (b) SmartAdP [LWL17].

3.4 Dimension 4 – Approaches for Shrinking the Information Space of Data

As a compact visual representation of the original data, an overview visualization aims to represent the most salient features of the data and thus needs to shrink the information space of the data to fit in the size of an overview display. The approaches include: (1) reducing the number of data elements; (2) reducing the visual representation; and (3) aggregating data elements.

Reducing the number of data elements is about selecting a limited number of data elements based on certain criteria and eliminating the redundant data elements. Commonly used techniques include filtering and feature extraction. An example can be found in Figure 3.4b.

Reducing the visual representation is about representing data elements in a succinct visual form (e.g., glyph, proxy, illustrative metaphor) to reduce the visual clutter. Examples can be found in Figure 3.4a and Figure 3.6c.

Aggregating data elements is about assembling data elements into a group of sets and representing every set as a unique visual representation. An example is the contingency wheel [AAMG12].

3.5 Dimension 5 – Techniques for Interactions and Detail Representations

This dimension deals with how to visually discriminate details of the data which are currently in focus from the overview display. Interaction with the overview display allows viewers to control: a. level of detail; b. extent of the data (spatial, values); and c. aspects of the data (attributes) in the detail representation. Techniques for interaction include: compare, sort, add variables, re-scale, re-visualize, filter, highlight, annotate, bookmark, aggregate, zoom and pan, and details on demand. Based on whether the overview information and detail information are jointly or separately displayed, detail representations fall into two major categories: Detail-In-Overview and Detail-Out-Of-Overview. Table 3.1 summarizes the techniques and provides some examples for each technique. These techniques will be elaborated with examples in the rest of Chapter 3.5.

3.5.1 Detail-In-Overview

The Detail-In-Overview paradigm broadly corresponds to the zooming, focus+context, and contextual cues in the contextual and detailed visualizations (as introduced in Chapter 1.3). It aims for a minimization of occlusions and visual cluttering as well as for a flexible visual filtering of the multi-variate data [CKB08]. The visualization integrates overview information and detail information into a single display so that both overview and detail are concurrently visible. In the

Table 3.1: Techniques and Examples of Detail Representation

Techniques			Examples
Details In Overview Display	Magnification	Physical	Magic Sphere [CMS94], Graph Lens [TAHS06], Magic Lens [BSP93]
		Semantic	OrthoZoom [AF06], ZAME [EDG08], OWL-VisMod [GTG11]
	Space Distortion		SignalLens [Kin10], Table Lens [RC94], Hyperbolic Browser [LR96], Radial Clustergram [ABF07], H3 Viewer [Mun98], 2D to 3D [CCF97]
	Visual Emphasis	Color	Two-Level Volume Rendering [HMBG00, 01], RTVR [MH01], IDTVDV [WYM08], SIMVIS [Dol07], LGE [KL06], GeoSpace [LI95]
		Opacity	Two-Level Volume Rendering [HMBG00, 01], RTVR [MH01], Activity Detection [OSBM14], Magic Lens [BSP93]
		Sharpness	Feature Enhancement [VKG05], SDOF [KMH01, 02]
		Style	Focus of Attention [VFSG06], Magic Volume Lens [WZMK05], FLOWLENS [GNBP11], ChronoLenses [ZCPB11]
Details Out Of Overview Display			ExoVis [TS02], LGE [KL06], Popout Prism [SWRG02], Prefuse [HCL05], Smooth Brushing [DH02], IFS [DGH03], Four-Level F+C [MKO08], O+D for DTI [ZCEV17], IDTVDV [WYM08], DOSA [VW14], CosMovis [HHB15], SIMVIS [Dol07]

visualization, overview information is represented attenuated or are omitted if they occlude detail information. The techniques of detail representation include magnification (physical or semantic), space distortion, and visual emphasis by color, opacity, sharpness, and style.

a. Magnification

Physical magnification increases the physical size of the region of interest. The magnified regions may occlude neighboring regions in the overview display. Figure 3.9a is an example of applying a lens to a dense graph (2D non-geometric data) to address its heavy clutter of edges [TAHS06]. In Figure 3.9b, physical zooming is applied to 3D data space where the voxels within the magnification lens are scaled.

Semantic magnification shows higher levels or aspects of information. An example is Zoomable Adjacency Matrix Explorer (ZAME) for exploring graphs at a scale of millions of nodes and edges. As shown in Figure 3.9c, the visualization allows analysts to explore a graph at many levels from an overview to the most detailed views [EDG08].

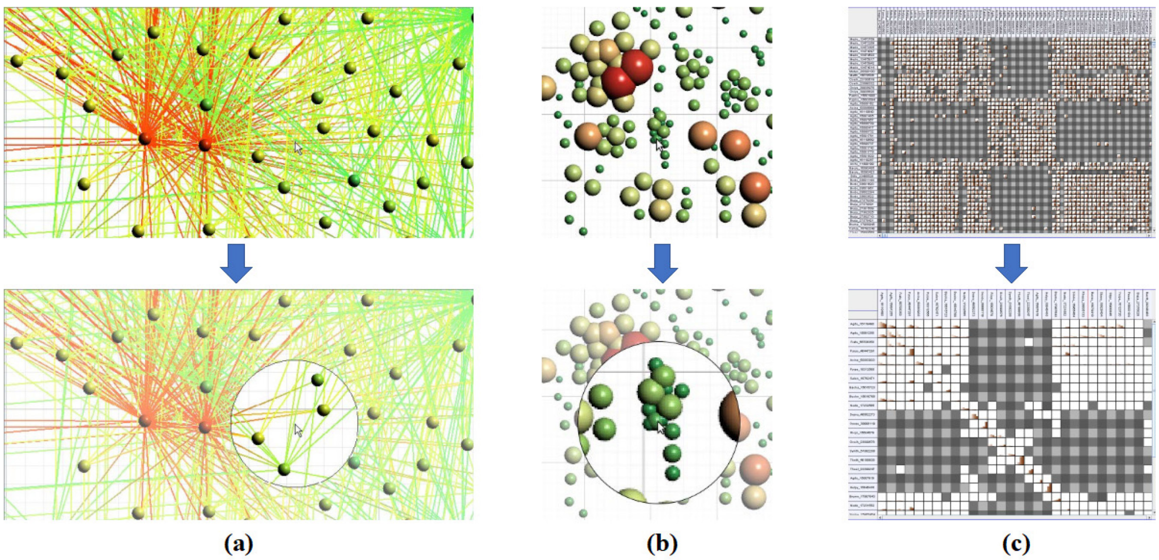


Figure 3.9: Detail representations using magnification. (a) A graph lens is applied to a dense graph visualization to identify which edges are connected to particular nodes in the focused area [TAHS06]. (b) A lens is applied to a 3D graph to magnify some nodes in the middle [TGK14]. (c) A protein-protein interaction dataset (100,000 nodes and 1,000,000 edges) is visualized using ZAME at two different levels of semantic zooming [EDG08].

b. Space Distortion

This mechanism uses more space for showing the focused area in more detail. Different from magnification in which the magnified area often occludes its neighboring regions, space distortion techniques display focus of interest seamlessly within its surrounding context. The less relevant peripheral items in the overview display are dropped or reduced in size. Examples of space distortion techniques include the fisheye lens, perspective wall, and interactive lenses in 1D (Figure 3.10a and b), 2D (Figure 3.10c), and 3D (Figure 3.10d and e).

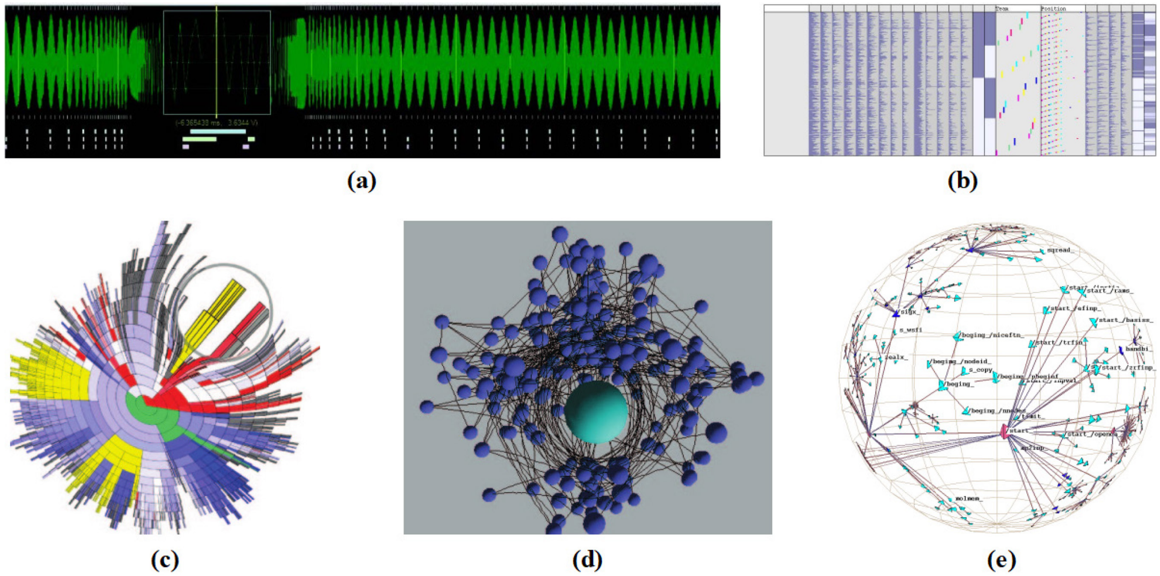


Figure 3.10: Detail representations using space distortion. (a) SignalLens visualizes an amplitude modulated electronic signal. The lens area exposes an anomalous glitch in the otherwise uniform oscillations [Kin10]. (b) Table Lens uses a fisheye technique that distorts tabular information to highlight focal area [RC94]. (c) In the Radial Clustergram, a lens is used to enlarge the area of interest while showing other portions of the graph in progressively less detail [ABF07]. (d) Space distortion technique is applied to a central node in a 3D graph [CCF97]. (e) The H3Viewer visualizes a graph in 3D hyperbolic space to see a large amount of context around a focus point [Mun98].

c. Visual Emphasis

Visual emphasis techniques emphasize the detail information in focus by using different graphics resources of color, opacity, image frequency, and style [Hau06]. Color-based techniques present focus in saturated/light colors while leave context in gray or dark colors (Figure 3.11a). Opacity-based techniques make focus rather opaque and context rather transparent (Figure 3.11b). Image frequency-based techniques use the difference between a sharp and blurred object depiction (Figure 3.11c). Style-based techniques differentiate the styles for context and focus. Usually, context in reduced style (Figure 3.11d).

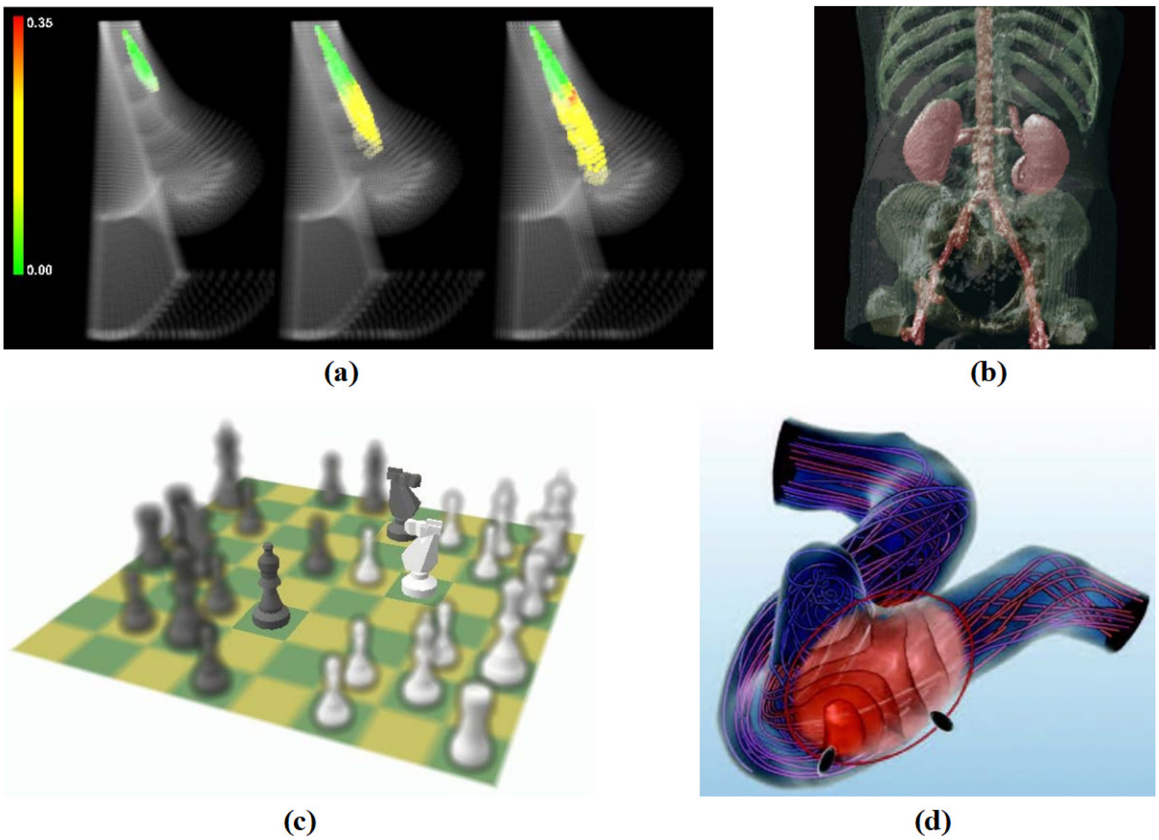


Figure 3.11: Detail representations using visual emphasis. (a) Color-based technique. SIMVIS shows an injection process in a diesel engine with the brushed data items highlighted in color which stands for the burnable amount of fuel-air mixture [Dol07]. (b) Opacity-based technique. Two-level volume rendering displays context information with low transparency while the inner structures with high transparency [HMBG00]. (c) Image frequency-based technique. A 3D SDOF chessboard with two active chess pieces highlighted while other chesses blurred [KMH02]. (d) Style-based technique. The FLOWLENS is applied to cerebral aneurysms [GNBP11].

3.5.2 Detail-Out-Of-Overview

This interaction paradigm is equivalent to overview+detail techniques in the contextual and detailed visualizations. The basic idea of this scheme is to display overview and detail information simultaneously and separately in different view ports or windows at different scales. The overview display and detail display can be organized as a coordinated pair or tiled into one window. Some typical examples can be found in Figure 3.12.

Compared with the Detail-In-Overview paradigm, Detail-Out-Of-Overview paradigm has a separate overview display which brings three major benefits. First, navigation is more efficient because users may navigate using the overview window rather than using the detail window [BW90]. Second, the overview window aids users in keeping track of their current position in the information space [PCS95]. The overview window itself might also give users task-relevant information, for example, by enabling users to read section titles from an overview of document [HF03]. Third, the overview gives users a feeling of control. A drawback of overview+detail interfaces is that the spatially indirect relation between overview and detail windows might bring some burden to users' memory and increase the time used for visual search and connection [HBP01]. In addition, an overview+detail interface either requires more screen space or sacrifices the size of the detail window for the overview window.

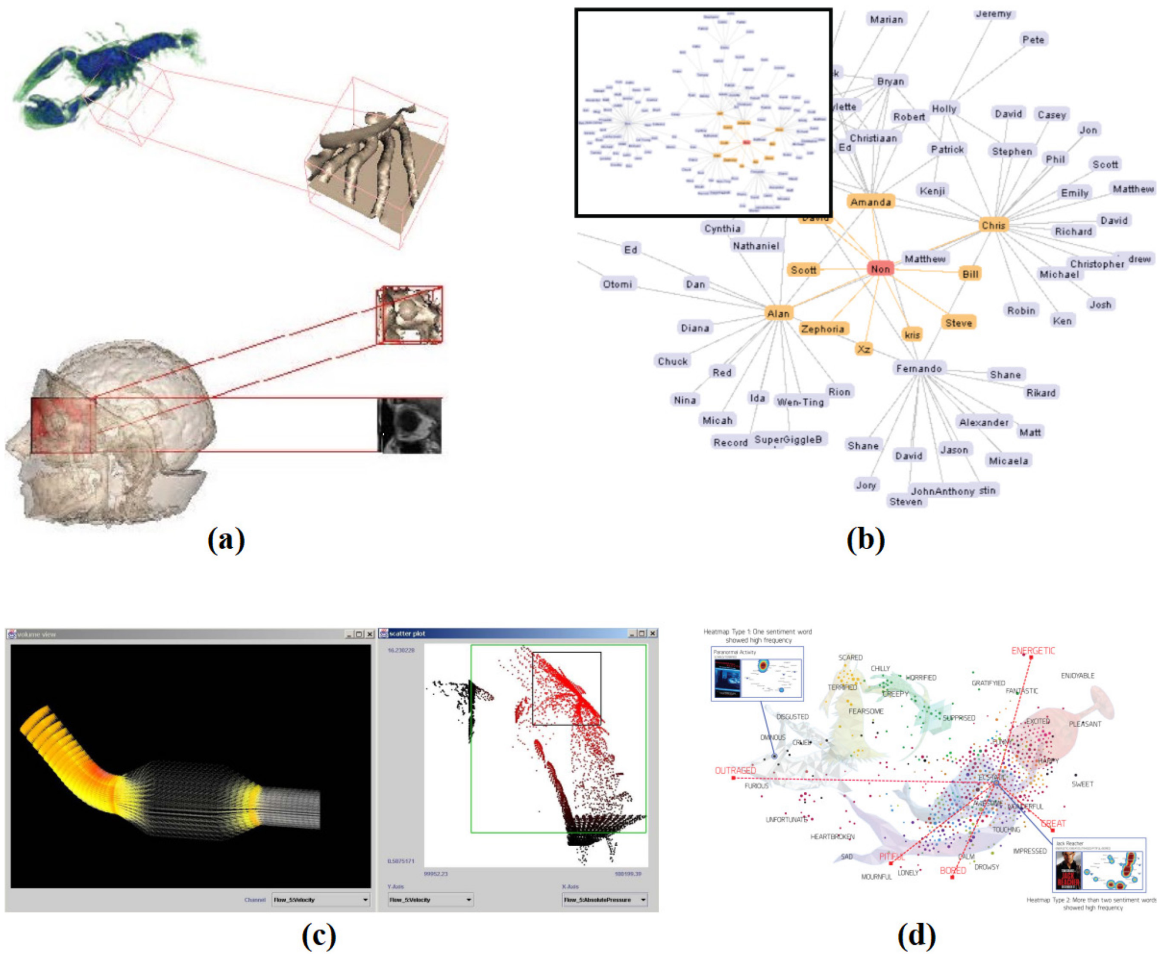


Figure 3.12: Examples of Detail-Out-Of-Overview. (a) The regions of interest are shown as callouts in the ExoVis [TS02]. (b) A force-directed layout with overview displayed in an inset separated from the main view [HCL05]. (c) Smooth Brushing and Linking: the contextual view is displayed on the right side as a scatter-plot where a non-binary DOI function was defined by smooth brushing a cluster of high velocity/high pressure data; the detailed view on the left side is a 3D view employing opacity/color modulation [DH02]. (d) A CosMovis constellation map of sentiment word-based movies. Movies can be selected from the words and displayed separately [HHB15].

Chapter 4: A Pipeline Model for The Development of Overview Visualizations

Based on the characterizations of overview visualizations discussed before and inspired by other visualization design models [CM97, CR98, PAK08, Mun09], a pipeline model is presented to split the development process of overview visualizations into five horizontal levels: (1) analyze domain problem and the related data; (2) identify design principle and the basics of overviews + characterize data; (3) design the overview display + transform data; (4) design the interactions and detail display + compute metadata; and (5) implement the overview visualization. The processes of visual design and data evolution in the visualization development are represented as two related pipelines in the model because they follow two distinct workflows from top to bottom but are mutually interdependent. The pipeline model considers human factors at each level as appropriate for human-centered visualization design. Figure 4.1 shows a flowchart of the proposed pipeline model. All of the five levels are connected by arrow in a way where the output from an upstream level above is input to the downstream level below. The two pipelines are connected by arrows at several levels as well. The rest of this chapter will elaborate each level of the model to give a guidance of the creation of overview visualizations.

4.1 Analyze Domain Problem and The Related Data

The goal of this level is to let a visualization developer get a comprehensive understanding of the domain problem and the related data. Starting with requirements analysis (or systems analysis), a visualization developer first needs to understand the problem to be solved, delimit its scope, find

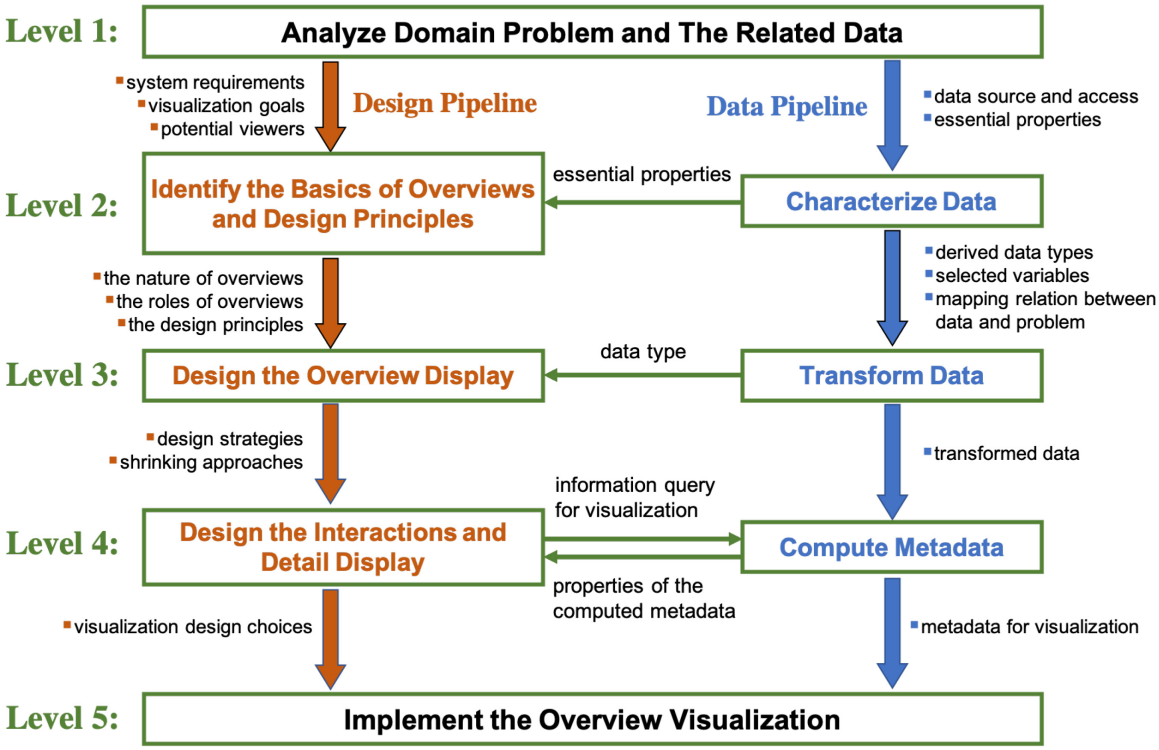


Figure 4.1: A pipeline model for the development of overview visualizations.

out how the problem is solved in the current practice and what are the limits of these solutions, and envision how the planned system might work. The methods of requirement analysis include observation, interviews, and surveys of potential users. The results are the identification of potential users and a set of system requirements which describes what the envisioned system-to-be is about in the vocabulary of domain. In a user-centered design, the results also include a set of user tasks [WL90, AS04, AES05, LPP06] to summarize the functional specifications and users' operations to the system-to-be. Followed by the requirement analysis, the designer needs to answer a set of visualization questions to map the problem from the domain scope to the visualization scope: (1) Why visualization is useful to solve the problem? (2) What type of visualization can be used to solve the problem? (3) What is the challenge of a visualization solution? The answering to these questions also helps designers derive a visualization goal.

The final operation at this level involves studying the data related to the domain problem. The visualization designer needs to find out the data source, access to the data, the essential properties (e.g., formats, size, variables) of the data, and the existing workflow of how the data is used to solve the domain problem.

The outputs to the next level include system requirements (or user tasks), potential viewers, visualization goal, and the basic understanding of the data.

4.2 Identify the Basics of Overview and Design Principles + Characterize Data

From this level, the development workflow is split into the visual design and data flow pipeline. In the visual design pipeline at this level, the operations involve identifying design principles and the basics of overviews which correspond to the first two dimensions in the characterization of overview visualization (see Chapter 3.1 and 3.2). The nature of overviews can be derived from the visualization goal set in the previous level. For instance, a phrase "to summarize the motion of" indicates an overview of change while "to give a gestalt view of the attitudes of voters" suggests an overview of content. Design principles, which form the basic rules of the visualization design, are identified from the system requirements or user analysis tasks from previous level. Sometimes, designers also need to consider users' desire or request. Based on the design principles, the roles of overviews can be derived to cover one or several design principles.

In the data flow pipeline at this level, data characterization involves finding out the right data type (e.g. time-varying volume data, categorical data, relational data) and variables so that the domain problem can be mapped to a data problem. For example, hotel booking records can be characterized as categorical data. Designers are recommended to first study the existing workflow of how the data is used to solve the domain problem before building the data model.

The outputs to the next level include the nature and roles of the overviews, the design principles of the visualization, the derived data type and selected variable, and the mapping relation between the domain problem and the data.

4.3 Design the Overview Display + Transform Data

At this level, the operations start with transforming the raw data into a derived data type based on the characterization of data in the previous level: quantitative data can be binned into a table of numbers where the columns contain quantitative, ordered, or categorical data; tabular data can be transformed into relational data with thresholding; a field of values or vectors at every point in a space can be transformed to a different field [Mun09].

The visual design pipeline at this level corresponds to the third and fourth dimensions in the characterization of overview visualization (see Chapter 3.3 and 3.4). The design strategy of an overview display usually accords with data type (e.g., abstract-context design strategy for an abstract data) but sometimes is affected by the roles of overviews. For example, in the case of time-centric design for time-varying volume data, the spatial information in the data are mapped to a non-spatial visual channel to highlight temporal variation. A contrary example is spatializing altitudes of an area (abstract data) to a map to add a spatial context. In general, designing shrinking approaches is more data dependent in scientific visualization and more application or user dependent in information visualization (see the discussions in Chapter 8).

The outputs to the next level include design strategies of overview displays, shrinking approaches, and the transform data.

4.4 Design the Interactions and Detail Display + Compute Metadata

The visual design pipeline at this level correspond to the fifth dimension in the characterization of overview visualization (see Chapter 3.5). The design of interaction paradigms usually needs to consider users' preferences and the characteristics of the overview display. For instance, users may prefer filtering visual components in the visualization by certain criteria instead of selecting or isolating these components from the visualization. On the other hand, designing a detail display is primarily upon designers. The pros and cons of displaying detail inside or outside overview are discussed in Chapter 3.5 along with different detail representation techniques (e.g., space distortion, visual emphasis by color).

In the data flow pipeline at this level, computing metadata involves creating analytical abstraction from the transformed data. The computed metadata will be used for visual encoding and mapped to the visualization. An example is a histogram of word frequencies in an article. The article is the original data and the frequencies of words are the metadata used for plotting the histogram. The metadata computation is usually determined by the derived data type from previous level and is affected by the visual design pipeline at this level (e.g., information query for visualization).

The outputs to the next level are visualization design choices determined at different levels of the visual design pipeline and the metadata for visualization generated from the data flow pipeline.

4.5 Implement the Overview Visualization

At this final level of the model, a visualization designer first needs to choose the operation system, programming language, and development tools, then determines the functionalities of the overview visualization based on the design choices and metadata. These issues of implementation are not unique to overview visualizations and are extensively discussed in the computer science literature.

Chapter 5: Introduction to The Case Studies

This thesis uses two case studies to demonstrate the creation process of overview visualizations and how overview visualizations provide an improved understanding of the data. In this chapter, I give a brief introduction to the case studies.

The first case study is illustrative visualization of Gulf Stream eddies. The data in this case study are a numerical simulation of the northwest Atlantic Ocean off the coast of North America generated by the Regional Ocean Modeling System (ROMS) [KC13]. Ocean scientists, who are the domain experts of this case study, would like to: (1) understand spatial relationships between eddy instances in a single time step; (2) understand the relationships of eddy instances over time (temporal relationships), and (3) observe the physical properties of single eddy instances. Combining the style and structure of an illustrative figure of ocean eddies with time varying feature-based visualization techniques, this work gives an overview visualization of ocean eddies in the data. The overview visualization provides multiple levels of representation of ocean eddies from a single eddy instance in a day or in an eddy path to multiple eddy paths in the context of ocean basin.

The second case study is visualizing job postings from Rutgers Career Portal (a job database). Understanding the relationship between jobs and college majors is of interest today in the higher education landscape but is not always straightforward. Students struggle to understand such connections. Many professors cannot provide clear links between what is being taught and opportunities that are available. In this work, we propose a novel visual exploration approach called JobViz that allows students (and professors) to explore job posting data in a more holistic and education-centric way. The JobViz gives an overview visualization of the whole job database and the relationship between majors, jobs, industries, and careers. The visualization design features the intuitiveness of node-link diagrams and the scalability of aggregation-based techniques.

Chapter 6: Case Study 1 - Illustrative Visualization of Mesoscale Ocean Eddies

The contents of this chapter were presented in the conference of EuroVis 2017 and were published in the journal of Computer Graphics and Forum (paper title "Illustrative Visualization of Mesoscale Ocean Eddies") [LSB17]. This paper serves as an example of creating overview visualizations for a 3D time-varying dataset (i.e., ocean eddies). The visualization in the teaser figure (Figure 6.1) gives an overview of the mesoscale ocean eddies and their motions in the Northwest Atlantic Ocean basin in the year 2006 and 2007.

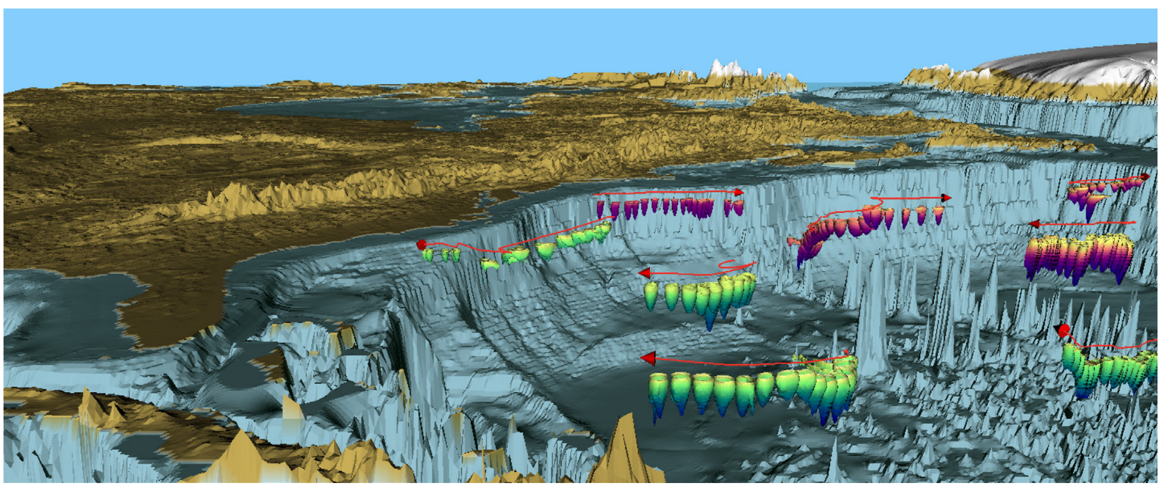


Figure 6.1: This image shows the Gulf Stream eddies moving through the Northwest Atlantic Ocean basin in 2006 and 2007. Each eddy instance is illustrated by the depiction of its color-coded isosurface. Colors are used to differentiate cyclonic (green) and anticyclonic (purple) eddy and convey variations in ocean temperature (warm=light; cold=dark). Directional lines are used to highlight the motion of eddies. This is an illustrative visualization using actual data in a manner similar to hand drawing illustrations (e.g., Figure 6.2)

Abstract

Feature-based time-varying volume visualization is combined with illustrative visualization to tell the story of how mesoscale ocean eddies form in the Gulf Stream and transport heat and nutrients across the ocean basin. The internal structure of these three-dimensional eddies and the kinematics with which they move are critical to a full understanding of ocean eddies. In this work, we apply a feature-based method to track instances of ocean eddies through the time steps of a high-resolution multidecadal regional ocean model and generate a series of eddy paths which reflect the life cycle of individual eddy instances. Based on the computed metadata, several important geometric and physical properties of eddy are computed. Illustrative visualization techniques, including visual effectiveness enhancement, focus+context, and smart visibility, are combined with the extracted volume features to explore eddy characteristics at different levels. An evaluation by domain experts indicates that combining our feature-based techniques with illustrative visualization techniques provides an insight into the role eddies play in ocean circulation. The domain experts expressed a preference for our methods over existing tools.

6.1 Background

As the Gulf Stream flows across the Atlantic Ocean, it develops meanders, which spin off to form high-vorticity features called eddies. These mesoscale ocean eddies play an important role in transporting heat, salt, and nutrients in the ocean, both horizontally and vertically. Three-dimensional computer modeling of ocean circulation has long been used to study the development of such eddies, but publications in oceanography typically show only simple two-dimensional plots of physical properties (e.g., temperature) on the surface and in vertical sections. This is partly because oceanographers are familiar with these views from observational studies (where available data may be only two-dimensional), and partly due to the limitations of the visualization tools that oceanographers typically use [WHP11]. While two-dimensional representations efficiently and simply illustrate the dominantly horizontal dynamics in the ocean, mesoscale eddies are places of vertical transport. The structure of mesoscale eddies along with their internal variation in temperature and salinity, and the migration and interaction of eddies over time and space, are all fundamentally three-dimensional. This motivates the development of fully three-dimensional visualization techniques that are appropriate to the analysis of mesoscale eddies in the ocean.

The use of visualization in the analysis of mesoscale ocean eddies addresses the need for insight into and illustration of the statistical population of eddies, the migration paths and potential transport by eddies, and the internal structure of eddies. Previous studies used visualization to provide a census of the eddies in a given dataset, visualize the spatial distribution of eddies [WHP11] [PWM13] [WPS16], and show the migration paths of eddies [ZM95] [PWM13]. Three-dimensional ray casting techniques and specialized colormaps were used to visualize the eddy structure and depth [SPA15]. Recent work on tracking eddies considers a variety of options in defining eddies [MAIS16]. Previous efforts at depicting eddy three-dimensional shape have primarily used abstractions (e.g., cylinders of eddy height in the Figure 4 and 5 of [WHP11]), reconstructed eddies from horizontal slices [ZM95], or just visualize all eddy instances in a data frame as a whole (e.g., Okubo-Weiss isosurface in the Figure 1 of [PWM13], and volume rendering of eddy kinetic energy in the Figure 1 of [SPA15]). The main challenges in this application domain include understanding spatial relationships between eddy instances in a single time step, understanding the relationships of eddy instances over time (temporal relationships), and being able to observe the physical properties of single eddy instances.

Idealized images of mesoscale ocean eddies have appeared in publications. For example, in Figure 6.2 one can see a hand drawn figure of mesoscale ocean eddies in the Pacific Ocean off the coast of Japan. The goal of this work is to combine the style and structure of an illustrative figure of ocean eddies (similar to Figure 6.2) with time varying feature-based visualization techniques to automatically produce illustrative visualizations with actual data. This combination will allow the entire dataset to be summarized into a single concise view but also enable structural understanding of the eddies and their dynamics. The dataset used in this work is a numerical simulation [KC13] of the northwest Atlantic Ocean off the coast of North America generated by the Regional Ocean Modeling System (ROMS) [SM03]. The first step used to process the data is feature extraction to identify potential eddies. After filtering for various criteria, eddies are tracked over a pre-specified time range. Several important properties (i.e. vorticity orientation, depth, etc.) of eddies are computed based on the metadata. Traditional volume visualizations of eddies are presented as isosurfaces colored by boundary values of a physical property such as temperature. These are combined with illustrative-based techniques including visual effectiveness enhancement, focus+context, and smart visibility. Both isolated and in context (i.e., in the ocean basin) displays of the eddy migration process serve to provide variable levels of detail and to tell the story of how mesoscale ocean eddies form in the Gulf Stream and transport heat and nutrients across the ocean basin. Careful color-coding of isosurfaces with visual effectiveness enhancement indicates eddy rotation and the value of a chosen physical property on the eddy boundary. Volume rendering of individual eddies hints at internal structure.

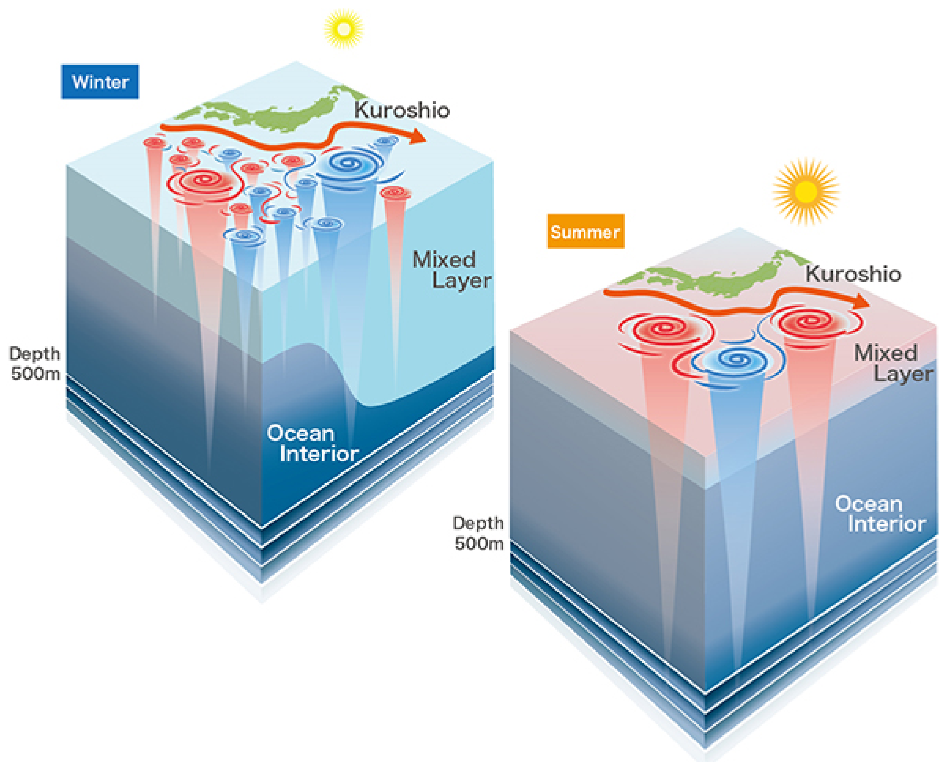


Figure 6.2: An artistic illustration of ocean eddies in the context of oceanic thermal structure. Image is courtesy of Japan Agency for Marine-Earth Science and Technology. [JAM]

The contribution of this work includes: (1) The combination of illustrative visualization with automated feature-based techniques to produce a figure summarizing a pre-specified time range of data; (2) enhanced feature extraction and tracking techniques especially for ocean eddies as independent entities to highlight their internal properties and individual movements and follow their migration paths; and (3) an evaluation by domain experts on the use of these visualization for real-time analysis and publications.

6.2 Related Work

Previous research related to the methodologies adopted in this work includes illustrative visualization, oceanographic visualization, and time varying volume visualization.

Illustrative visualization focuses on generating more abstract or expressive visualizations typically inspired by works from artists and illustrators [RBGV08]. Hand-drawn images have been used for centuries to explain and depict natural phenomena and the main goal of illustrative visualization is to mimic these images by computing and abstracting the data automatically. The resulting visualization is a pictorial abstraction of the dataset (or datasets) as opposed to a direct rendering. Illustrative visualization of time varying data is especially helpful as it is difficult to convey motion without animation, and watching many time-frames is time consuming. In time-varying volume data, the process of illustrative visualization often relates to using metadata to enhance perception, manage visibility and provide focus in renderings of the data or its events (e.g., [JR05, JR08, JCRS09, LS08, HMCM09, BWF10]). A variety of techniques including visual effectiveness enhancement (primarily focused on non-photorealistic rendering (NPR) [RE01]), visibility management (or smart visibility) [VG05], focus+context [VKG05] [Hau06], and interactive visual storytelling [WH07] [MLF12] are useful for scientific visualization. Many of these techniques are summarized in [BCP12]. The motion of features over time can be depicted using a combination of tracking, variable transfer functions, and artistic renderings. Compositing of multiple time steps combined with selective colormaps to control visibility has been used effectively to convey the evolution of volume features [JCRS09] [HMCM09]. In many cases, compositing in combination with selective visibility management can accurately convey the motion or evolution of flow features.

Understanding the motion of eddies and tracking that motion is of increasing interest to the oceanographic and meteorological communities. Zhu and Moorhead [ZM95] reconstructed the three-dimensional shape of ocean eddies from eddy boundaries identified on a vertical sequence of horizontal slices and tracked eddy instances over time by calculating their correlation probabilities between successive time instances. In visualizing the results, they placed the reconstructed three-dimensional structures of eddies in an ocean bathymetry of the Pacific Ocean off the coast of Japan and used track lines in a two-dimensional map to report the motion of ocean eddies over time. Grant et al. [GEO02] visualized the vertical structure and the life cycle of large anticyclonic eddies associated with the loop current by showing the raising and sinking of the thermocline associated with eddy genesis and decay. Contextualizing the eddy location and height within an ocean basin and its thermocline, Williams et al. [WHP11] presented a skeletonized view in which each eddy instance is simplified as a cylinder coded with different colors indicating the rotation direction. The height of the skeletal lines indicated the eddy vertical extent. Petersen et al. [PWM13] tracked eddies from a high-resolution simulation of the global ocean and provided a global census of the eddies. The authors presented the three-dimensional shape and spatial distribution of the eddies as isosurfaces of the Okubo-Weiss parameter. Samsel et al. [SPA15] developed specialized colormaps and a new three-dimensional ray casting technique to visualize the structural depth of ocean currents and eddies with high fidelity. Woodring et al. [WPS16] developed an in-situ workflow for eddy analysis in a high-resolution ocean climate model. Matsuoka et al. [MAIS16] explore two-dimensional methods of defining and tracking eddies with the context of ocean currents. Our work extends some of their tracking techniques to three-dimensions.

Visualization of time-varying volume data has long been of interest to the visualization community. Methods generally include compressing data to make it more manageable [SCM99] [MS00]; preselecting transfer functions for rendering [PLB01] [JKM01]; using hardware-accelerated rendering [ECS00] [LMC02]; and computing features and tracking them [SSZC94, SW97, RPS01, JSW03, JS06, CJR07, MM09, BWP10, CMN13]. Amongst these methods, automated feature computation techniques using domain knowledge can significantly reduce the amount of data that needs to be visualized by focusing on just those regions of interest. In this work, we adopt the feature-based techniques that depend on feature extraction and tracking. The process known as feature extraction and tracking consists of identifying interesting regions or coherent structures of a dataset (i.e., features), automatically searching for these features, and tracking these features as they move and interact from one time step to the next. Various feature tracking models have been proposed to track time-varying features and to visualize their evolution over time (referenced above). Most of the algorithms extract features from each time step using a variant of segmentation or isosurfacing, and then try to match extracted features from one time step to the next using volume overlap, pattern matching or feature-vector matching. One can also track groups of features [OWS12] and identify actions and activities [OSBM14].

6.3 Data and Approach

6.3.1 ROMS Data

The ocean data used for this study is from a multi-decadal regional ocean model simulation [KC13]. The simulation was performed using the Regional Ocean Modeling System (ROMS) [SM03], which solves the incompressible, hydrostatic, primitive equations with a nonlinear free surface. It employs orthogonal curvilinear coordinates in the horizontal and a generalized topography following coordinate in the vertical. The simulation domain covers the major path of the Gulf Stream through the northwest Atlantic (Figure 6.3).

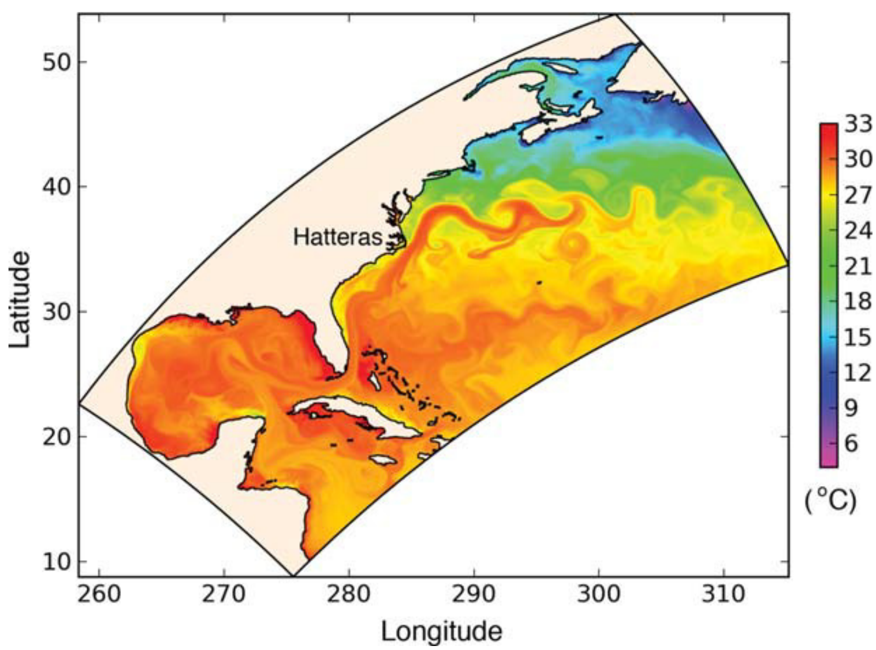


Figure 6.3: ROMS domain and a representative model output of the daily averaged sea surface temperature. [KC13]

In this study, the chosen dataset to process covers the most recent two years (2006 and 2007) in in this ROMS simulation with 730 daily-average NetCDF files (approximately 1.67 GB per file).

Each data frame has 760×360 grid points with a horizontal spacing of 7 km and 40 vertical terrain-following levels stretched toward the surface so as to resolve the surface boundary layers. Following the ROMS convention (https://www.myroms.org/wiki/Vertical_S-coordinate), the vertical coordinate transformation and stretching functions are types 2 and 4, respectively, with the surface and bottom control parameters θ_s=7, θ_b=2 and the stretching parameter=250. The data contains 56 variables such as temperature, salinity, and velocity vectors. Figure 6.3 [KC13] demonstrates the typical model output of the daily mean sea surface temperature (SST) for a given day. The mesoscale meanders and eddies of the GS are clearly seen.

6.3.2 Approach Overview

Our approach combines time varying feature-based techniques with illustrative visualization. Investigating eddies as independent features highlights their internal properties and individual movements. The availability of the full voxel set for each eddy feature enables the use of a variety of traditional volume visualization techniques (e.g., isosurfacing and volume rendering) and enhancements such as designed colormaps that can depict the physical properties (e.g., temperature, salinity, and velocity) along the eddy boundary or within the eddy. The focus of our method is on illustrating the changes in eddy shape or physical properties during migration in the context of the broader ocean or population of eddies, which is crucial but is not always apparent from standard visualization techniques. Several visual effectiveness enhancement techniques are applied in our work. Cylinders are used to simplified eddy instances and highlight their positions and vertical movements, directional lines and opacities to indicate eddy motion direction, and ocean basin bathymetry to add context. We found that the illustrative flow visualization techniques such as visibility management (or smart visibility) [VG05], interactive visual storytelling [WH07] [MLF12], and context+focus [VKG05] [Hau06] give us sufficient artistic freedom to depict eddy kinematics in an expressive way, while simultaneously providing ocean scientists an insight into and enabling them to naturally interact with the data. Guided by these techniques, we place the visualized eddies in the context of time (compositing of time steps) and space (the ocean basin and the location of the eddy along its migration path) and connect all of the proposed visualizations with each other to provide variable levels of detail and to tell a full visual story. A framework of the proposed method can be found in Figure 6.4.

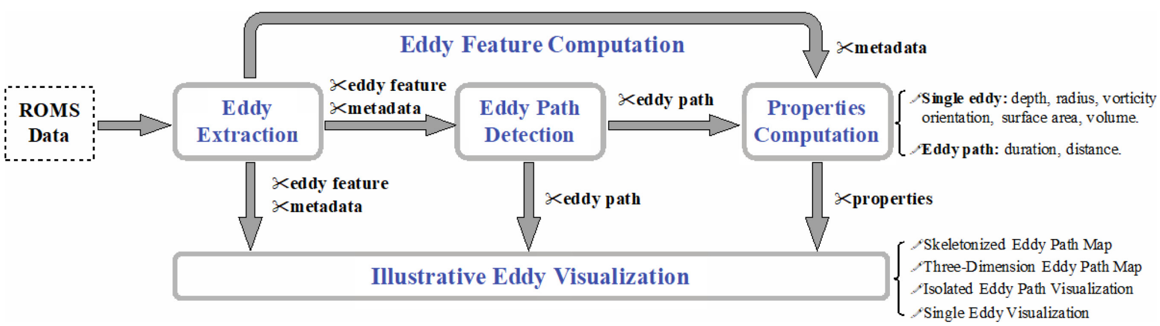


Figure 6.4: The proposed method has two phases: (1) Eddy feature computation includes three steps: eddy extraction, eddy path detection, and properties computation. All of these provide data that is used to construct the illustrative eddy visualization. (2) Illustrative eddy visualization.

6.4 Eddy Feature Computation

6.4.1 Eddy Feature Extraction

In the past few decades, several methods have been developed to detect eddies from satellite measurements and numerical simulations (see [MAIS16]). In this work, we use the Okubo-Weiss

(OW) criterion to detect eddy from ROMS data, because this method has shown considerable success and is the standard vortex criterion within the oceanographic community. The OW parameter [Oku70] [Wei91], which describes the relative dominance of deformation with respect to rotation of the flow, is one of the most widely used physical quantities to detect eddies. In general, this parameter divides the ocean into regions dominated by vorticity, regions dominated by strain or deformation, and a background where neither effect is dominant. Regions dominated by vorticity are potentially eddies. Eddies can be identified where the calculated OW values exceed a given threshold. Mathematically, the OW is defined as:

$$OW = s_n^2 + s_s^2 - \omega^2 \qquad (1)$$

where s_n and s_s are the normal and shear components of strain respectively, and ω is the relative vorticity of the flow. They are different combinations of partial differentiations of the horizontal velocities (u and v) in the x and y directions.

$$s_n = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \qquad (2)$$

$$s_s = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \qquad (3)$$

$$\omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \qquad (4)$$

We follow the above formulas and compute the OW parameter in every grid point. In Figure 6.5, an OW parameter field is shown by plotting the OW parameter in a ROMS data frame. The

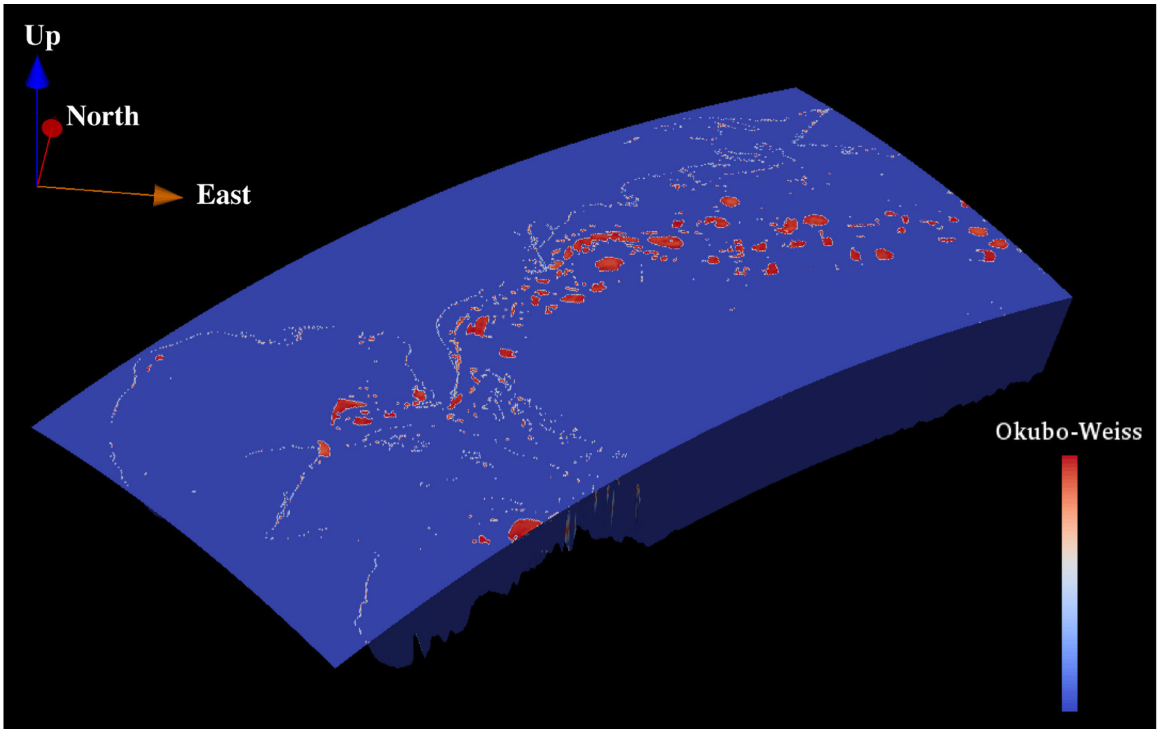


Figure 6.5: The Okubo-Weiss parameter is plotted on a data frame.

ROMS uses a geographic coordinate system with the spatial coverage of the data frame from 258.34° to 315.19° in longitude, 8.78° to 53.83° in latitude, and 0 meters to 5500 meters in altitude. Beyond computing the OW parameter, we also pull five three-dimensional variables from ROMS: temperature, salinity, velocity in longitude direction (u), velocity in latitude direction (v), and vorticity sign, and save them with the computed OW parameter in every data point.

Given a positive threshold value, the OW parameter allows partitioning of the ocean data into strain dominated and vorticity dominated domains. To select only vorticity dominated features, we followed the convention in the ocean community and set the threshold as $0.2\sigma_{ow}$, where σ_{ow} is the spatial standard deviation of an OW parameter field. A region-growing algorithm [SSZC94] [SW97] with a threshold of $0.2\sigma_{ow}$ is applied to the OW parameter field in every data frame. Note that the threshold value changes in each data frame as the OW parameter depends on time. Connected components of points with OW values below the threshold are extracted from the data as vorticity-dominated features which are potential eddies. Each feature contains the values of the OW parameter and the five variables (temperature, salinity, u, v, and vorticity sign) at each point. Spatial properties of each extracted eddy feature, such as volume (number of grid points), centroid, and bounding box, etc. are also computed in the process of extraction.

Though the OW parameter can identify eddies both at and below the surface of the ocean, it also identifies high-vorticity features that should not be considered as eddies, particularly meanders in the Gulf Stream and other strong currents [WHP11]. To filter out non-eddy high-vorticity features, we combine the OW criterion with some geometrical constraints introduced in [KC13] and identify eddies from all of the extracted features ("eddy candidate") in three steps: Step 1. Detect the local minima and maxima of sea surface height (SSH), as well as the local minima of the OW parameter and sea surface velocity magnitude; Step 2. Combine the minima of the OW parameter, velocity field, and the minima of SSH to determine cyclonic eddy centers; Step 3. Combine the minima of the OW parameter, velocity field, and the maxima of SSH to determine anticyclonic eddy centers. Features that fail this geometrical constraint test are filtered out. Here, cyclonic and anticyclonic are vorticity orientations. In the north hemisphere, cyclonic eddy spins counter-clockwise and anticyclonic eddy spins clockwise.

6.4.2 Eddy Path Detection

Since an eddy is considered as a temporal phenomenon, the complete description must include all of its instances along the time axis and the associated properties [ZM95]. Tracking is used to determine which eddy instances in consecutive time steps belong to the same eddy. We call this spatial-temporal process of an eddy migration over its life circle an eddy path. To detect eddy paths from the eddy extraction results, all of the identified eddy features are correlated over time with a volume-based feature tracking algorithm [SW97]. During the migration process, an eddy feature can split, propagate, and merge (see also [MAIS16]). The feature tracking algorithm computes various attributes including tracking history of features (correspondence list), position changes, and any other value/attribute that is a function of two consecutive time steps.

Tracking is only the first step in detecting eddy path. With numerous eddy features spanning hundreds of time steps, scientists need an automatic procedure to isolate each eddy path with their states indicating their interactions in the evolution. In our work, we utilized a state graph-based event detection framework introduced in [OSBM14] to capture these eddy paths. In this method, the state graph is represented as an enhanced Petri Net formalism and encapsulates all of the states a modeled event includes. As shown in Figure 6.6: the eddy path detection model uses five circles to represent five states (S1, S2, S3, S4, and S5) that any eddy may go through from its birth to death: birth (S1), propagate (S2), merge (S3), split (S4), and death (S5); the gray rectangles sitting between the circles define the transition conditions or actions between the states; the directed arcs define the connections between the states.

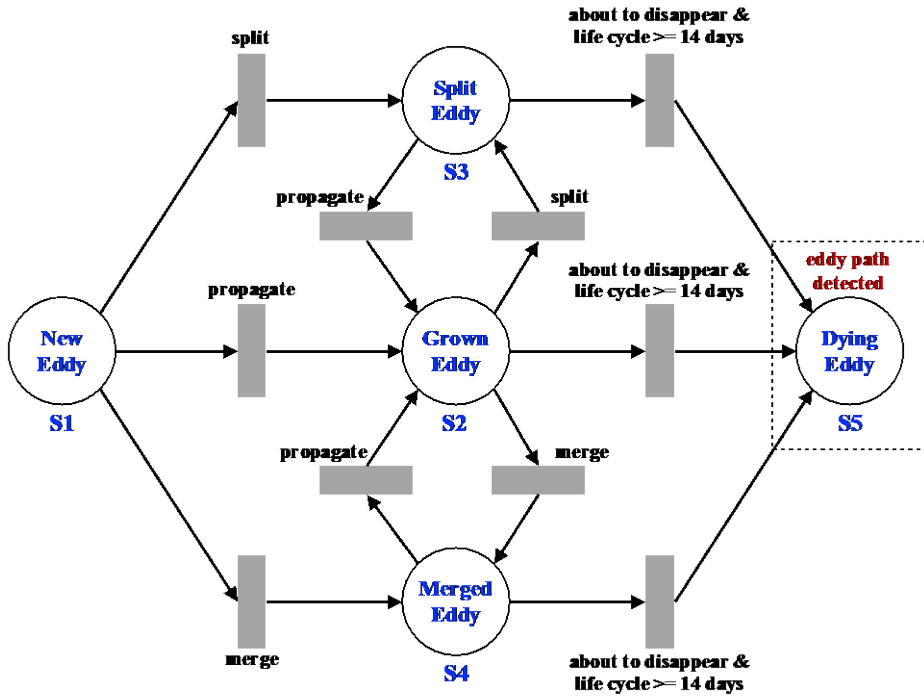


Figure 6.6: Eddy path detection model. An eddy that lasts at least fifteen days will be detected. Each output eddy path includes eddy features' IDs and their corresponding state IDs in each time step.

Execution of the eddy path detection algorithm starts from reading the tracking history and feature metadata and is done by means of tokens which represent instances of individual eddy features in a time step. The location of a token in the model describes the current status (new birth, propagated, split, merged, or about to die) of the eddy feature that token represents. The output file includes the participating eddies' IDs and their corresponding state IDs (1, 2, 3, 4, 5) in each time step for each detected eddy path.

We tested the eddy path detection model for reachability (all other eddy states can be reached from the birth state) with different state and transition conditions. In our testing, we designed an automatic error detection mechanism to handle the possible errors. We modified our eddy path detection model to allow tokens to move twice within a timestep and forbid tokens moving from the final state to any other states. The oceanographers who have collaborated in this project gave us much help in designing and modifying the model. Based on their feedback, we added an eddy's minimum life span criterion of fifteen days, which means any detected eddies that lived less than fifteen days were excluded in the final output

6.4.3 Properties Calculation

Using the metadata computed in the Chapter 6.4.1 and 6.4.2, our method has a capability to calculate several important geometric and physical properties of eddy.

Depth. The depth of an eddy defines how deep an eddy penetrates into ocean from its top. This information can be computed by finding the distance between the upper X-Y plane and lower X-Y plane in the bounding box.

Radius. The eddy radius is defined as $R_e = \sqrt{A/\Pi}$ Where R_e is the equivalent radius of a circle with the same area (A) enclosed by the bounding box of eddy feature.

Surface Area. The surface area of an eddy is computed by the Maximum Unit Normal Component (MUNC) algorithm [LEA92] as follow:

$\Delta a_i = \frac{\Delta x \Delta y}{|n_{z_i}|}$ if n_{z_i} is the maximum unit normal component; (5)

$\Delta a_i = \frac{\Delta x \Delta z}{|n_{y_i}|}$ if n_{y_i} is the maximum unit normal component; (6)

$\Delta a_i = \frac{\Delta y \Delta z}{|n_{x_i}|}$ if n_{x_i} is the maximum unit normal component. (7)

Where Δx, Δy, and Δz are the dimensions of the *i*th grid; $n_{x_i}, n_{y_i}, and\ n_{z_i}$ are the unit normal vector components. Here Δx and Δy are firstly calculated in kilometers from the longitude/latitude points.

$$Surface\ Area = \sum_i^N \Delta a_i \qquad (8)$$

Volume. The volume of an eddy is computed by the Divergence Theorem Algorithm (DTA) [LEA92] as follow:

$$Volume = k_x \sum_i^N \left(x_i n_{x_i} \Delta a_i\right) + k_y \sum_i^N \left(y_i n_{y_i} \Delta a_i\right) + k_z \sum_i^N \left(z_i n_{z_i} \Delta a_i\right) \quad (9)$$

where x, y, and z are the coordinates; k_x, k_y, and k_z are coefficients whose sum is equal to one; n_{x_i}, n_{y_i}, and n_{z_i} are the unit normal vector components; and Δa_i is determined by equation (8).

Vorticity sign. This sign defines the orientation of an eddy's spin. It is positive if an eddy feature is cyclonic, and negative otherwise. This property is computed in identifying eddies from extracted features (Chapter 6.4.1).

Duration. The duration of an eddy is equal to the number of timesteps (here is days) in its evolution path.

Migration Distance. The migration distance of an eddy is equal to the total distance between eddy features in the eddy path.

6.5 Illustrative Eddy Visualization

In this section, illustrative visualizations are combined with standard visualization techniques such as isosurfacing and volume rendering to explore eddy characteristics and motions at four different levels: (1) *Skeletonized Eddy Path Map* where eddy instances are represented as simple glyphs (cylinders) and the path, a life span composited set of instances, is placed in the context of the ocean basin, Gulf Stream, and coastline to provide a summary view of eddy distribution with eddy vertical transfer highlighted. (2) *Three-dimensional Eddy Path Map* which replaces the cylinders in the *Skeletonized Eddy Path Map* with color-coded isosurface and places them in a more detailed context. (3) *Isolated Eddy Path Visualization* which illustrates the motion of an isolated eddy in space with a directional line and opacity to convey direction. (4) *Single eddy visualization* where an eddy feature can be color-coded as an isosurface or volume rendered to convey eddy rotation direction and map the temperature, velocity or other physical property variation inside and on the surface of it. A standalone GUI has been developed to implement these proposed visualizations (Figure 6.7). The GUI loads the selected eddy path sequences and provides a *Skeletonized Eddy Path Map* at first. Users can then switch to *Three-dimensional Eddy Path Map*. The visualized eddy paths in both maps are filterable by duration. An eddy path can be isolated and visualized in a new window where a single eddy instance is further investigated. The calculated eddy properties in Chapter 6.3 are provided along with the visualizations in the GUI.

6.5.1 Skeletonized Eddy Path Map

The Skeletonized Eddy Path Map is designed to give a qualitative overview of the vertical extent and spatial (horizontal) distribution of ocean eddies within time-varying volume data such as results from numerical simulations (e.g., ROMS). It is faster and permits observation of a larger portion of the dataset than the *Three-dimensional Eddy Path Map*. Figure 6.7 shows an example of our use of skeletonized eddies to depict eddy paths in the northwest Atlantic off the United States coastline in 2006 and 2007. The context of the visualization is provided by the combination of ocean basin bathymetry (from the ROMS model domain) and a flat land surface indicating the east coast of the US from which we created a terrain elevation surface using Delaunay Triangulation [Rup95]. The

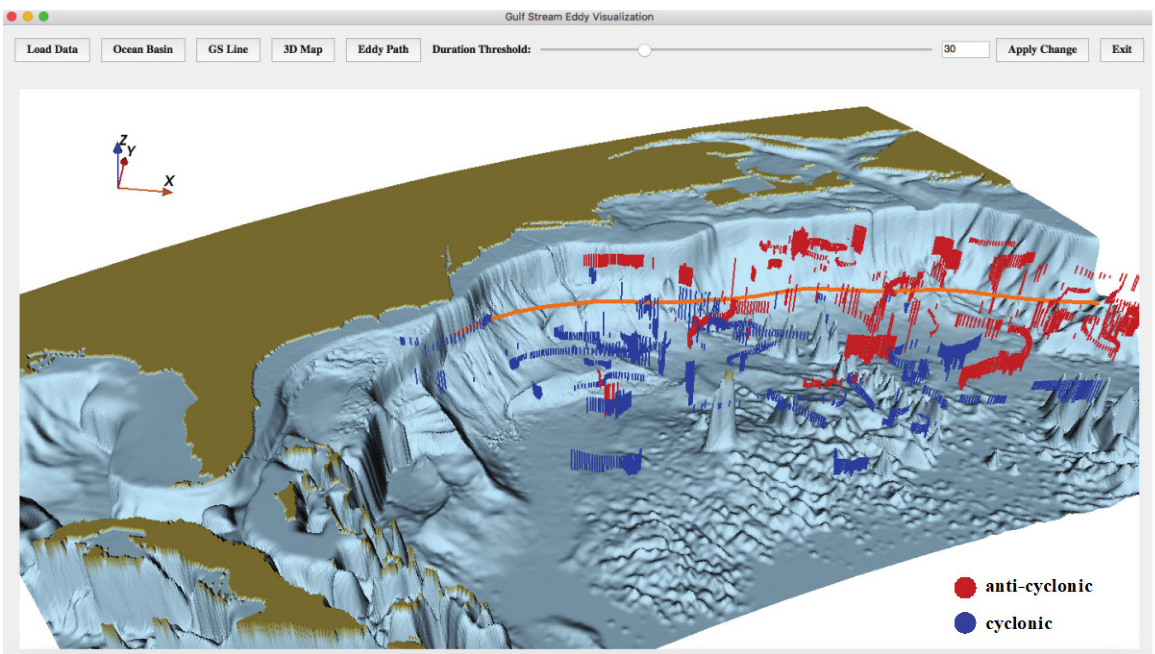


Figure 6.7: The GUI is showing a Skeletonized Eddy Path Map in 2006 and 2007. The eddy instances in the eddy paths lasting at least fifteen days are represented as simple glyphs (cylinders) and placed in a context of the ocean basin, Gulf Stream (orange line), and coastline. Color is used to differentiate cyclonic eddies (blue) and anticyclonic eddies (red).

typical Gulf Stream is drawn as an orange line on the sea surface. In this map, each eddy instance is represented as a cylinder with its top and bottom at the same depth as the highest point and the lowest point in the eddy, similar to the eddy skeletonized view in [WHP11]. Based on the computed eddy vorticity orientation, either blue or red color is assigned to the cylinders to differentiate cyclonic (blue) and anticyclonic (red) eddies. In Figure 6.7, a minimum lifetime cutoff of thirty days is set and only eddies that meet that criteria are displayed. The *Skeletonized Eddy Path Map* visualizes the migration path of each month-long eddy path as a time composited sequence of cylinders within the ocean context. As we can see from the figure, cyclonic eddies are more likely to appear to the south of the Gulf Stream line while anticyclonic eddies are just the opposite. Also, most eddies extend less than 2000 meters below the sea surface (the deepest point of the ocean basin is 5500 meters).

6.5.2 Three-dimensional Eddy Path Map

Although the Skeletonized Eddy Path Map presents an efficient and succinct overview of the distribution and migration of ocean eddies, some important information such as the size, shape, and physical properties of eddies and the interactions between eddies is missing. To give a more informative summary of eddy migration, a Three-dimensional Eddy Path Map is proposed (see examples in Figures 6.1 and 6.8). This visualization replaces the eddy skeletons in the Skeletonized Eddy Path Map with isosurfaces of eddy features whose surfaces are color-coded by ocean temperature. Two different colormaps are applied to maintain the differentiation between the cyclonic (green colormap) and anticyclonic (purple colormap) eddies. A more refined context is provided by the use of a higher resolution elevation dataset from National Oceanic and Atmospheric Administration [NOA] in combination with actual topography (based on publicly available global elevation data) in the generation of a joint terrain surface. To avoid excessive occlusion between eddy features within an eddy path, we sample every eddy path at a sampling rate of one feature for every three days. We then added a directional line above each eddy path to provide a sense of continuity and convey the direction of eddy motion. In Figure 6.8, the physical

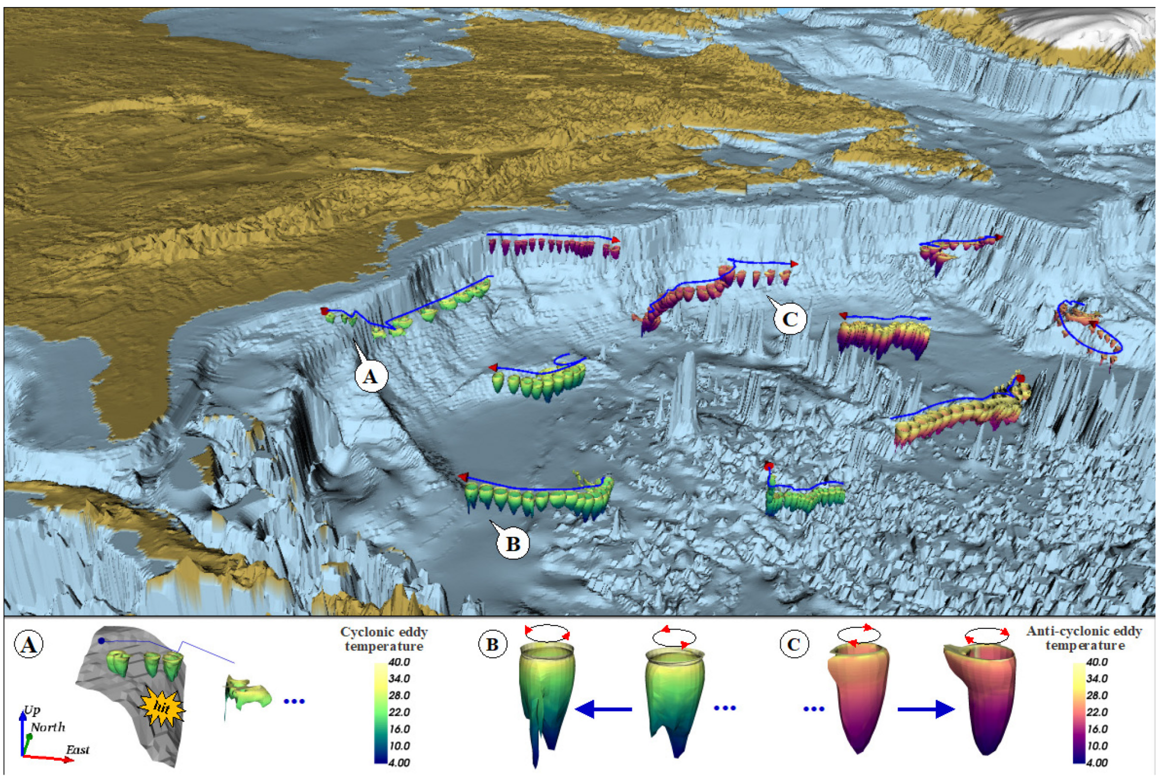


Figure 6.8: This three-dimensional Eddy Path Map shows the eddy paths that last at least fifty days in 2006 and 2007. The physical and kinematic changes in an eddy over its life cycle are illustrated by the depiction of its three-dimensional color-coded isosurface in successive time steps. Directional lines above each eddy path highlight eddy motion. Colors convey variations in ocean temperature. Three typical scenarios are observed from the map: A. An eddy propagates onto the shelf, then bumps it before splitting. B. A cyclonic eddy propagates to the west. C. An anticyclonic eddy propagates to the east.

tapering of the eddy and the penetration of the thermocline by the eddy are clearly illustrated. The three insets along the bottom of the figure animate three typical scenarios of eddy dynamics observed in the map: an eddy propagates onto the shelf, then bumps the cliff before splitting into several “smaller eddies”; a cyclonic eddy propagates westwards; and an anticyclonic eddy propagates eastwards.

6.5.3 Isolated Eddy Path Visualization

Sometimes an oceanographer wants to focus on the details of a single eddy path. The Isolated Eddy Path Visualization presents every time step of an isolated eddy path as an isosurface color-coded by temperature, or other user-specified physical property (e.g. salinity, velocity), without any distracting context. This visualization highlights the physical changes an eddy goes through in space and time. To avoid occlusion by successive eddies, the distance between visualized eddy features is exaggerated by a factor of ten relative to the spatial scaling of the eddy isosurface. Following the opacity modulation techniques in [JR05], eddy features in older time steps are shown as a faded representation. Directional lines also convey eddy motion during the structural and temporal changes. Figure 6.9 presents four examples, two cyclonic and two anticyclonic eddy paths, of the proposed Isolated Eddy Path Visualization.

6.5.4 Individual Eddy Visualization

The Individual Eddy Visualization helps scientists observe the internal structure and the variation of physical properties both inside and on the surface of a picked eddy. Examples of a cyclonic eddy

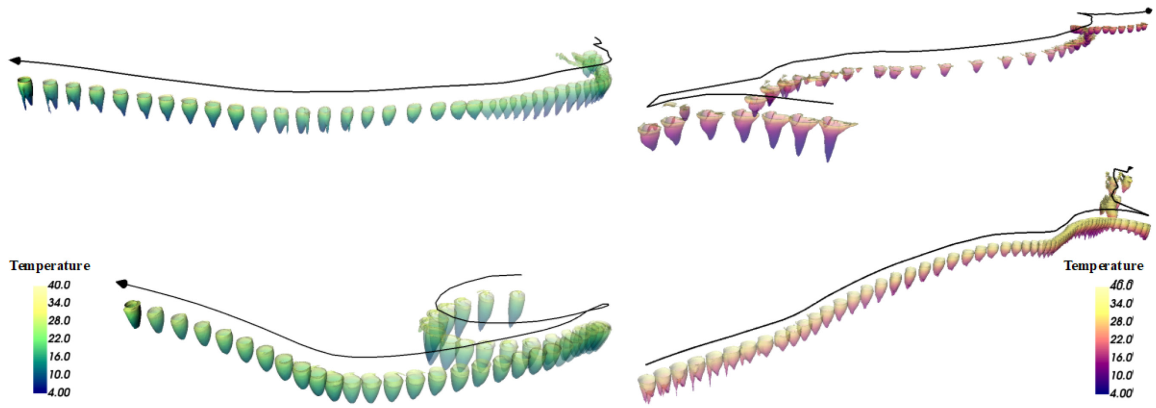


Figure 6.9: Two cyclonic eddy paths (left) and two anticyclonic eddy paths (right) are presented as Isolated Eddy Path Visualization. Transparency and directional lines are used to give a sense of motion. Each eddy feature (instance) is plotted as isosurface color-coded by temperature variation on its surface.

and an anticyclonic eddy (selected from two eddy paths in Figure 6.9) are visualized in Figure 6.10 and 6.11. The variation of a selected physical property that is either temperature, salinity, or horizontal velocity on each eddy surface is visualized as a color-coded isosurface. Different colormaps are consistently used with different properties to assist the scientist’s memory and recognition. To gain insight into the internal structure by considering the internal variation of physical properties, the two eddy features are clipped with a cutting plane and are rendered with standard ray-casting volume rendering techniques. From Figure 6.10, we can see that temperature drops both inside and on the surface of the two eddies as they extent deeper into ocean, and that the anticyclonic eddy in the upper row of the figure has a warm core (i.e. the center is warmer than outside) while the cyclonic eddy in the bottom row of the figure has a cold core (i.e. the center is colder than outside). Moreover, we observe that temperature decreases gradually within the thermocline from 200 meters to 1000 meters in depth, while almost keeping constant below the

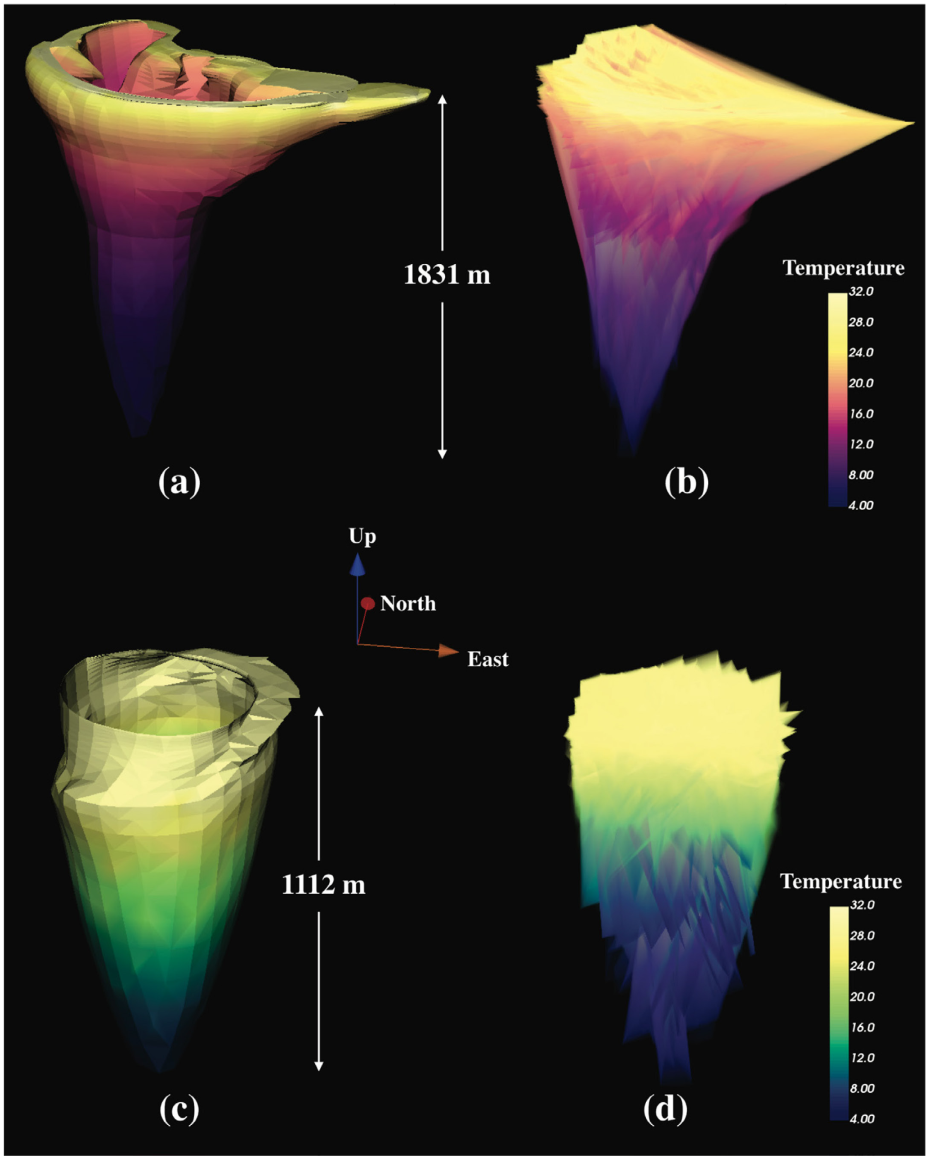


Figure 6.10: Individual Eddy Visualization of temperature for an anticyclonic eddy (a, b) and a cyclonic eddy (c, d). In (a) and (c), temperature variation on the surface are color coded in isosurface of the two eddies. In (b) and (d), eddies are first clipped by a cutting plane, then their internal structures are shown by volume rendering indicating the variation of temperature inside.

thermocline. In Figure 6.11, horizontal velocity of the anticyclonic eddy and salinity of the cyclonic eddy are visualized showing that both salinity and horizontal velocity decrease as the two eddies go deeper into the ocean.

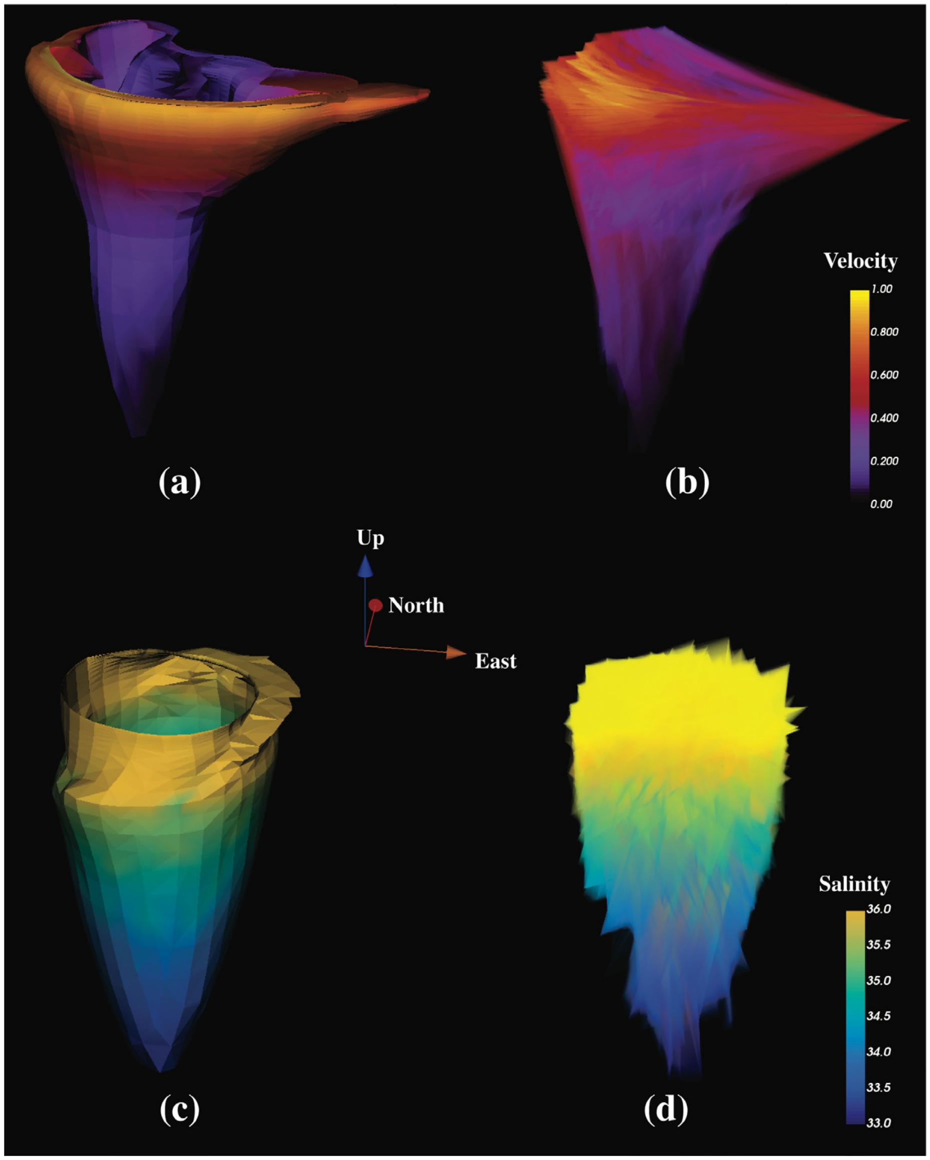


Figure 6.11: Individual Eddy Visualization of horizontal velocity for an anticyclonic eddy (a, b) and salinity for a cyclonic eddy (c, d).

6.6 Expert Evaluation

Our method and visualization results were reviewed by the three domain experts at Rutgers University who have collaborated in this project: Karen Bemis, who specializes in applying three-dimensional visualization to volcanology and marine geophysics; Enrique Curchitser and Dujuan Kang, the physical oceanographers who provided the ROMS data. Our expert evaluation was designed to obtain subjective and critical feedback from these experts.

6.6.1 Methodology

We first gave a presentation to introduce our methodology and show the eddy visualization results to the experts. We then invited them to use our GUI to explore and interact with the proposed visualizations. Finally, they gave their both written and oral comments on the figures of eddy visualization (Figure 6.1, 7, 8, 9, 10, 11 in this chapter) to evaluate the effectiveness of the proposed illustrative eddy visualization in addressing the main challenges of this application domain: whether the method was able to convey information about the ROMS data, namely, the spatial relationships between eddy instances in a single time step, the temporal relationships of eddy instances over time, and the internal structure and physical properties of single eddy instances.

6.6.2 Expert Feedback

This subsection first describes the feedback from the experts on individual visualizations and then summarizes the overall impressions. The proposed Skeletonized Eddy Path Map shown in Figure 6.7 was particularly liked by an expert as, “This figure provides a clear overview of the evolutions and distributions of the long-lived cyclonic and anticyclonic eddies in a year. The ocean bottom topography is vividly illustrated. The duration threshold can be adjusted to show the life cycles of eddies with different life times.” Furthermore, the experts said that this technique has potential in summarizing the data in different time-ranges, but could be even more useful if combined with some on-the-fly statistical analysis.

All of the experts preferred the proposed *Three-Dimensional Eddy Path Map* in Figure 6.1 and 6.8 over the existing two-dimensional eddy road map (e.g., Figure 1 in [DBSL07] and Figure 6 in [ZM95]). The figures helped them understand the different life histories of anti-cyclonic and cyclonic eddies and how these are affected by interactions with the ocean basin: "These figures provide insights into the eddies' on-shore and cross-stream transport of the momentum and energy and both figures tell a story of what happens to Gulf Stream eddies." Furthermore, one of the experts thought showing the life path of detected eddies in the context provides a "nice interesting way to show if eddies generated in a particular region are more likely to end up on the shelf." The experts were surprised at how few eddies actually approach the shelf and that most anti-cyclonic eddies move away from the shelf.

The *Isolated Eddy Path Visualization* in Figure 6.9 was liked by all of the three experts, because: "This figure illustrates trajectories of the long-lived cyclonic and anticyclonic eddies in 3D. Visualizing eddy path as a series of 3D entities provides more details than traditional 2D vertical profile view and tracking lines. The color-coded isosurface clearly conveys the temperature variation on eddy surface." The experts were also amazed to see that the eddies almost invariably taper to a point at depth, "The tendency to narrow with depth has implications about the geostrophic and shear conditions of the ocean and how they change with depth." They called this visualization as "an effective and succinct way to observe the temporal, spatial, and physical changes of a small number of eddies."

For Figure 6.10 and 6.11, experts were interested in the capacity of the proposed Individual Eddy Visualization in showing the profiles of temperature, salinity, and vertical velocity in the eddies chosen from the context, “These two figures give me a close view to eddies. The technique has the potential to improve the understanding of inner-eddy structures or distributions of different properties, such as temperature, salinity, or biogeochemical variables.” One issue that did come up was the difficulty of “seeing inside” the volume rendered results in these figures. Alternatives were suggested including adding slice planes and volume cuts (similar to Figure 5 and 7 in [CLTE11]) which show vertical profiles of eddies using temperature anomaly.

Experts were also pleased to be provided with an interactive tool that allows them to choose their view. While statistics are needed for publications, the sliding bar on the GUI (see figure 6.7 for one example in the GUI) provided immediate feedback on the differences between eddies based on eddy lifespan. The interactivity provided by the GUI allowed the experts to ask questions: “I finally got to ask ‘can I see more details on that anti-cyclonic eddy that moves for weeks in the opposite direction to most anti-cyclonic eddies?’ and be able to get a direct answer.” The connection between eddies in the regional views and their isolated instances provides powerful query capability as the length of simulation and hence the number of long-lived eddies increases.

The expert evaluation process gave us much more qualitative information regarding how our method can help develop a better understanding of the ocean circulation. In general, we found that the experts thought our techniques are effective in helping them understand eddy properties and motions. And they liked the GUI we developed that allowed for interactive exploration. The novelty of our method is also favored by all of the three experts one of whom wrote in a follow-up email:

"It constitutes a significant upgrade to state of the art tools for visualizing the ROMS data and that it is novel with respect to the integrated display of paths and shapes in the context of ocean basins." Although the color-code isosurface of eddies is favored by the experts, the traditional two-dimensional techniques in showing eddy internal structure were preferred over the cut-away volume rendering techniques in Figure 6.10 and 6.11. We think that this is partially due to the limitations of volume rendering and partially because of their familiarity with the two-dimensional visualization technique which allows them to interpret the values in a chosen slice. The scientists are currently writing up a publication based on the finding of the tracked eddy data. The experts were satisfied with the overall processing and visualization performance. The combined processing time from feature extraction to illustrative visualization (order of hours) is much less than the ROMS simulation run time (order of days to months). In the interactive viewing tool, which uses precomputed processing, the response to changing the sliding bar on eddy lifespan is instantaneous.

6.7 Conclusion

Visualization plays a key role in developing scientific insight into ocean circulation. The goal of this work is to combine the hand drawn figurative depiction of mesoscale ocean eddies (e.g., similar to the Figure 6.2) with time varying feature-based visualization techniques to automatically produce illustrative visualizations from a high-resolution regional ocean model. This combination will allow the entire dataset to be summarized into a single concise view but also enable structural understanding of the eddies and their dynamics. Investigating eddies as independent features highlights their internal properties and individual movements with a flexibility to provide illustrations of various scope (including ocean basin scale, eddy migration path scale, and individual eddy scale). The proposed illustrative visualization method effectively conveys the spatial relationships between eddy instances, temporal relationships of eddy instances over time, and the variation of physical properties both inside and on the surface of eddy instances. Variable levels of detail provided by the proposed visualizations help scientists understand how mesoscale ocean eddies form in the Gulf Stream and transport heat and nutrients across the ocean basin. An expert evaluation was conducted to identify the effectiveness of our techniques. The experts' feedback indicated that our method gives an insight into the role eddies play in ocean circulation.

Given this initial success, we intend to extend the capabilities of our illustrative eddy visualization techniques. As ocean dynamics and its interactions with the ocean ecosystem are inherently multivariate problems, we plan to develop multi-variable visualizations that convey a greater sense of the interactions. In addition, the transport capacity of eddies is related to the differences between the internal properties of eddies and the properties of the surrounding ocean, so we plan to develop modern visual methods that convey that same sense of differential seen in two-dimensional cross-sections.

Acknowledgement

We thank the Department of Environmental Sciences at Rutgers University for the ROMS data and the domain experts who gave us feedback. This work has been funded by US National Science Foundation (NSF) Grants OCE-1049088. We would also like to thank Dr. Min Chen for his valuable suggestions.

Chapter 7: Case Study 2 – Interactive Visualization of Job Posting Data

The contents of this chapter were presented in the conference of IEEE VIS 2018 at Berlin, Germany and were published in the journal of IEEE Computer Graphics and Application (paper title "Application Driven-Design: Help Students Understand Employment and See The "Big Picture") [LSB18]. This paper is a case study of creating overview visualizations for an abstract dataset (i.e., job postings in a university job database).

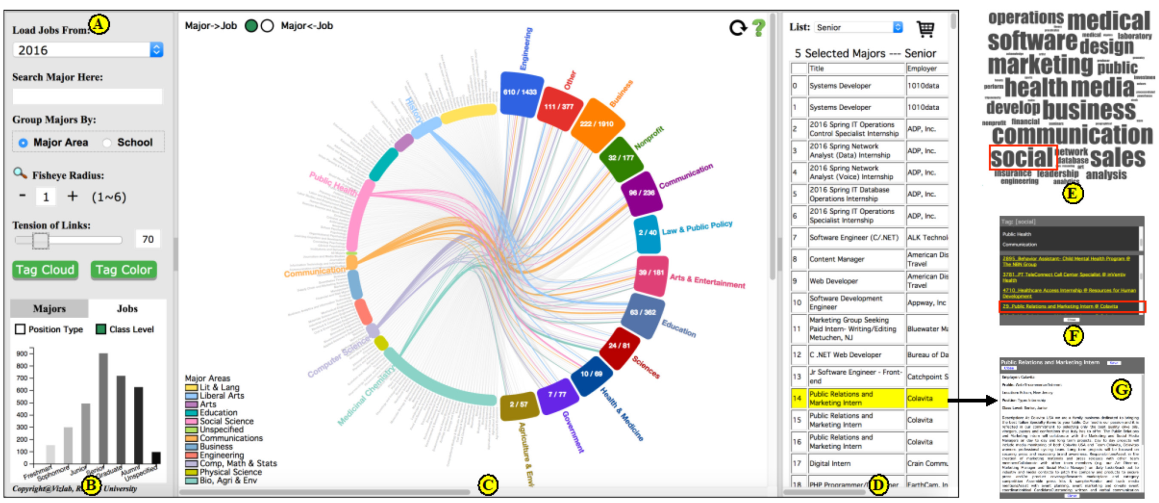


Figure 7.1: The user interface contains three modules. The left module has (a) a control panel, and (b) an inset of a statistical view. The center module (c) visualizes the job posting database as a circular layout. Majors are on the left side, and career sectors are on the right. A user can select majors or career sectors to see their relationships in a Relation View. Here a Relation View visualizes the links between five selected majors and career fields containing job postings requesting at least one of these majors. (d) The right module lists job postings. In the Tag Cloud View (e), the keywords of job postings that are listed in (d) are displayed as tag cloud. Each term in the tag cloud is clickable and is linked to a list of job postings as in (f). A particular job posting can be identified from (d) or (f).

Abstract

Understanding how college preparation relates to employment prospects is crucial in positioning a student for successful post-college employment. Yet few tools for job seeking or job market assessment utilize visualizations. Visualizing job postings and helping students gain a better sense of the job market require both the representation of hundreds of majors and thousands of jobs and the display of the intangible relationships between them. To address these challenges, we present JobViz, a novel visualization designed for interactive exploration of job postings from a university career portal and for use with career counseling. The application-driven design combines the intuitiveness of node-link diagrams and the scalability of aggregation-based techniques to provide an overview of the entire database and to allow users to individualize and explore the data. We demonstrate the effectiveness of JobViz with a case study and two questionnaire-based user studies.

7.1 Background

In today's "competency-based education" model, universities are being pressured to make sure students are graduating with the skills they need to be directly employed. Students may or may not have already decided upon a major, but are nevertheless interested in their potential job prospects. Therefore, assisting students to understand the links between majors and job opportunities and providing students with effective career counseling is becoming more and more important at universities. Where do students go to get information on available jobs, desired majors, and necessary skills? Many students utilize the counselors at the university career services or their academic advisors, who then have them use the career portal or commercial web-based job search engines (e.g., Indeed, Monster, etc.). These tools rely solely on text-based querying and a linear presentation of results, and so they do not give novice users a general overview and understanding of the data. Most students are forced to query the list-based job portals using keywords they may have thought of or majors that they are considering. The students then discuss the results with their advisors. However, they cannot get an overview or “gestalt” view of all of the data.

A visualization system that presents the entire university job posting database and enables data exploration and analysis in a more holistic, interactive, and education-centric way would be a valuable asset to career counseling. With such a system, students can get a better sense of the demands in the workplace, make more informed decisions about their major choices and class selections, and adjust their job seeking approach to maximize success and satisfaction.

Major design challenges center on the need to (a) present an accurate summary of relationships between the large number of majors and even larger number of job postings, (b) counter the many preconceived notions students have about majors and employment, and (c) overcome both presentation bias and selection bias that can readily bias the novice user (students). Presentation bias relates to the ease of overlooking portions of a long list based upon where in a list a particular term is placed. Selection bias refers to when users search only for familiar items thereby potentially missing out on a big picture view of the data. The ultimate goal for JobViz is to use visualization as part of a career counseling process.

The job posting data can be characterize as categorical data. Multivariate categorical data contain a series of variables whose values comprise a set of discrete categories [FS00]. The discrete nature of categorical data inherently requires different methods from those commonly used for quantitative data. Previous techniques (see Chapter 7.2) have been developed to visualize the associations between categorical data variables. However, the techniques are less effective with large numbers of categories and data entries. Displaying the intangible relationships between categorical variables (e.g., majors and jobs) is also a challenge.

To address these challenges, we designed JobViz, an interactive visualization system for university job postings (Figure 7.1). Combining the intuitiveness of node-link diagrams and the scalability of aggregation-based techniques, JobViz presents a visual overview of the relationships between majors and jobs by treating four variables (i.e., major, job title, position type, and class level) as interconnected sets. The proposed visualization features the simplicity and familiarity of a circular layout, hierarchical edge bundles, and tag clouds and provides both an overview of the database and the capacity for interactive queries to retrieve contextualized details. With this layout, we also tried to remove any inherent bias in the data to enable students to look at and understand all of the data.

In this chapter, we discuss our design process and show how an overview visualization can help students gain a better sense of the job market. We have performed two user studies of our visualization: one for general feedback and the other as part of a career development class where it is being utilized as a component of the curriculum.

7.2 Related Work

The problem of displaying associations between categorical variables can be treated as visualizing relations between sets. Categorical variables take discrete non-ordered values comprising a set of categories. Data entries are classified into the categories. Visualizing sets and their relations have been approached in various ways including Euler diagrams, matrix-based techniques, scatterplot-based techniques, node-link diagrams, and aggregation-based techniques [AMA14]. The proposed visualization approach relates most closely to node-link diagrams and aggregation-based techniques.

Node-link diagrams model the membership relations between elements and sets as edges of a bipartite graph [AMA14]. Bipartite graphs can be represented by different layouts. For example, Jigsaw [SGL08] places the elements (categories) of two sets (variables) in two lists parallel to each other and links these elements by lines to represent the overlap. Parallel coordinates [ID90] extends the idea of parallel list visualization to n-dimensional (n>=2) categorical data by placing n vertical lists (axes) side by side. Each n-dimensional data entry is represented by a polyline which intersects the parallel axes. Circos [KSB09] arranges elements (categories) of sets (variables) around a circular layout and connects them by ribbons of varying thickness. The ribbon thickness encodes the number of data entries that fall in both elements.

Aggregation-based techniques combine multiple elements belonging to a specific overlap between sets into a single visual object. They often summarize categorical variables into a contingency table before analysis. A typical example is Mosaic Display [FS00] which represents overlapping elements of sets (cells of a contingency table) as a collection of rectangular “tiles” whose areas are proportional to the strength of overlap (cell frequencies). Parallel Sets [KBH06] improves parallel coordinates [ID90] by representing sets as boxes with size proportional to the number of elements and substituting frequency based ribbons across sets for polylines. Contingency

Wheel [AAMG12] is designed for visualizing asymmetrically large 2-way contingency tables where the number of rows greatly exceeds that of columns. Contingency Wheel treats each column (a category of a low cardinality variable) as a set and visualizes them as circular sectors. The cells along each column are treated as elements of sets and are represented in histograms to show the breakdown. Arcs show overlaps between pairs of sets.

There are many existing job posting sites both free (e.g., Monster [Monster], Indeed [Indeed]) and subscription-based (e.g., Burning Glass [BGT], Emsi [Emsi]). Domain professionals have used various information visualization techniques (e.g., bubble charts [Gap12], tag cloud [Big13], and circular display [UCB14]) to create an overview of employment data. Many of these techniques play an important role in explaining reports, accompanying studies, and assisting decisions making. While some have visualizations relating to queried data, none provide visual interfaces to the data.

JobViz combines the advantages of node-link diagrams and aggregation-based techniques to constitute a multi-level overview+detail exploration interface. Jobs are aggregated into career sectors for scalability, but the visualization emphasizes each major element and career sector as individual objects around a circular layout and explicitly represents their intersections by spline curves. A tag cloud displays the summarization of keywords of job postings.

7.3 System Design

The overarching goal is to design a visualization that provides a holistic educative overview of a database of job postings. This initial goal was refined and expanded during the design process. The tool is targeted towards undergraduate students who are novice users that do not have a background in the labor and career field. However, to reach the students the tool gets distributed through career counselors and educators. While counselors are conscious of design choices, they are also aware of pitfalls and biased assumptions that students may have. The design must take this into consideration as well. Occasionally these influences suggest conflicting design choices. Driven by the application domain, our design followed the nested model [Mun09] at different stages.

7.3.1 Data Characterization

The job posting data used in this work is from the Rutgers University career portal. While this is specific to one university, many universities use the same system [Symplicity] with the same data format. The career portal contains about 10,000 job postings at any one time. Companies fill out a questionnaire when posting a new job including information such as Employer, Job Title, Position Type, Majors, Class Level, Posted Date, Location, and Job Description. The system expects companies to choose which major they think is most applicable to the job postings. However, companies can also choose a set of majors or have the option of choosing "All Majors" (i.e., indicating they have no preference for a particular major). This data is similar to job posting data on public sites except that major information is displayed and these positions tend to be entry-level.

We pull data from the career portal and save it in a CSV spreadsheet in which each row records a job posting. Data updates can be done through processing table entries. Our system design mainly focused on the visual representation of five categorical variables (i.e., majors, job titles, position types, class levels, and job descriptions) and the connections between them.

7.3.2 Tasks Abstraction

The potential end users of the system-to-be include college students, university career counselors, and faculty. As part of our design process, we worked closely with the career counselors at Rutgers Career Services and interviewed with four professors (two from computer science, one from electrical engineering, and one from environmental science) and seven students to collect their requirements.

We found that all of the potential users would like to see a big picture of the relationships between college majors and job postings. Students, our main audience, also want to know what types of jobs are available and what majors and skills are most requested by employers in the job market. Career counselors needed intuitive tools and interactive presentation techniques to help students understand more about the data. The professors were especially interested in learning how the course materials they teach relates to the employment sectors. Based on the requirement analysis, we compiled a list of user analysis tasks as follow.

T1 *What does the entire database of job postings look like?*

T2 *How are majors and jobs related to each other?*

- ***a.*** *What jobs require a particular major?*
- ***b.*** *What majors are appropriate for a particular job?*

T3 *Can more detail be shown for a particular major or career field?*

T4 *Can the actual job posting be displayed?*

T5 *What credentials are required by a particular job posting or set of job postings (e.g., a career field or industry sector)?*

These tasks summarize the topics, concerns, and sentiments expressed in the interviews and were later confirmed through the design process.

7.3.3 Design Principles and Visual Encoding

Through the task analysis, we identified the following key design principles that form the basis of the system: ***DP1.*** Show an overview of all data relationships including connections between majors and jobs (***T1, 2***); ***DP2.*** Provide interactive query and access to all levels of data details including actual job postings (***T2, 3, 4***); ***DP3.*** Visually summarize the requested credentials for categories or groups of jobs (***T5***); ***DP4.*** Reduce the information bias (as requested by the career counselors and discussed in Chapter 7.7).

The visual encoding was iteratively guided by the above data characterization and design principles over the course of two years of collaboration with students, counselors, and experts in visual design. In each design iteration, several design choices were proposed and explored. We also demonstrated our prototypes to the users and invited them to evaluate the prototypes for effectiveness, to identify the limitations, and to provide design suggestions for improvement.

The initial design started with identifying the crucial variables. For four of the five categorical variables, we grouped data entries into categories. Each variable was treated as a set and its categories as elements in a set. Among the set visualization techniques [AMA14], node-link diagrams were judged suitable for visualizing these sets because node-link diagrams visually emphasize the elements as individual objects and explicitly represent their connections. For the final visual component, we decided to summarize the keywords of job postings by creating a tag cloud [HR08] of a collection of skills, industry knowledge, and other facets from the job descriptions. This design decision was driven by the fact that the representation of a tag cloud is compact and draws the eye towards the largest, and presumably most important items [HR08]. Based on these considerations, we designed two bipartite layouts in which the first (Figure 7.2a) places the elements in two parallel axes (similar to parallel coordinates [ID90]), and the second (Figure 7.2b) arranges elements in a detached circular layout (similar to Circos [KSB09]). The tag cloud of keywords was placed in the center

We presented these two design choices of node-link diagrams with pre-attentive visual components and properties (Figure 7.2) to several students and a professor with a background in visual design. They all favored the circular layout over the parallel layout. One advantage of a circular layout is that it is compact display. It makes effective use of screen real estate as it occupies a large circular area in the middle of the screen while reserving the sides and corners for other

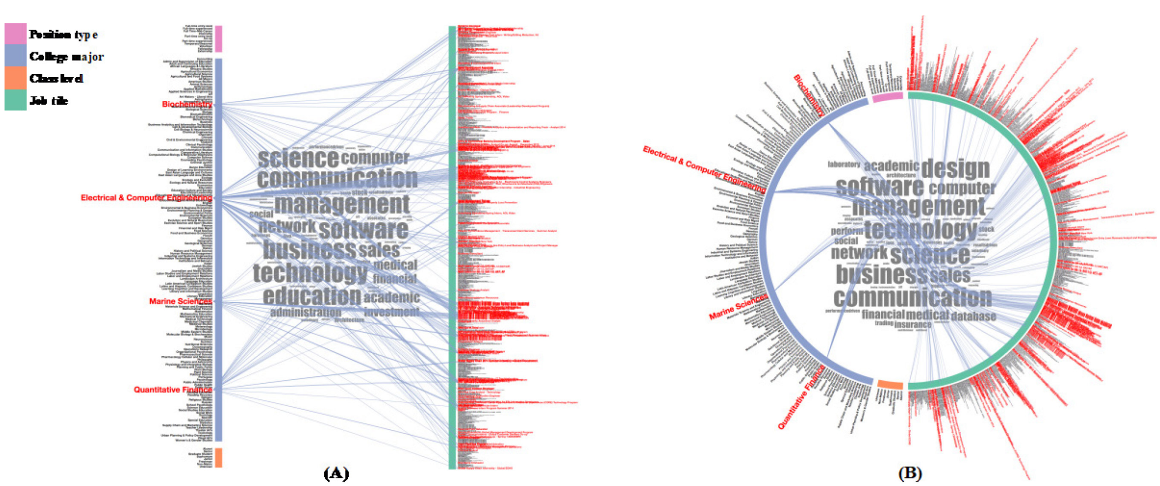


Figure 7.2: Two design choices of the node-link diagram for comparison in our initial design. (A) places elements in two parallel coordinates and (B) arranges elements in a circular layout. (B) was preferred.

purposes (e.g., legend and buttons as shown in Figure 7.1c). Furthermore, resolution varies linearly and increases with radial position within the circle. This makes the center of the circle ideal for compactly displaying low resolution summary information (i.e., links and tag cloud) which the reader can then follow outward to explore the data in greater detail. The third advantage of a circular layout is its neutrality which was requested by the career counselors. By placing visual elements at equal distances from each other and from the center of the drawing, none is given a privileged position. The first prototype of JobViz was based on the visualization design in Figure 7.2b. The circumference of the circular layout was divided into four separate sections for majors, job titles, class levels, and position types. Each job title in the visualization is linked to a job posting. The center of the circular layout was left for displaying major-job links and the tag cloud. We represented the major-job links as spline curves based on the Hierarchical Edge Bundles [Hol06], because they were more compact than straight lines. The first prototype was demonstrated to a career counselor and dozens of students in a workshop. The audiences all liked the components of the visualization: the detached circular layout, the major-job links, and the tag cloud. However, they thought the visualization became too cluttered when there were over 1000 job titles listed.

In developing the second prototype, we realized that we need not visualize majors, job titles, class levels, and position types in the same way. The last two variables can be used as criteria to filter jobs rather than displayed visually. Majors are a primary part of the message so we kept the explicit listing of majors from the first prototype, but clustered majors by major areas. Segments with unique colors distinguish majors from different major areas (Figure 7.1c). We stepped back to our initial design choice of set visualization approaches and recognized that aggregation-based techniques have higher scalability in handling the large number of elements and so are better for representing jobs. Driven by the audiences' feedback and the design of Contingency Wheel [AAMG12], we grouped job titles into twelve career sectors [CF] and represented them as detached circular sectors with unique colors. Numerical labels on each career sector indicate the number of jobs they contain. Unlike Contingency Wheel which displays overlaps between sectors, the career sectors here are independent from each other (i.e., no job postings belong to more than one career sector). The relationships between majors and career sectors are presented as links which connect majors on the left semi-circle to career sectors on the right semi-circle.

As shown in Figure 7.3, we designed three patterns of career sectors. The first pattern used equal-size sectors and attached three bars to each sector to represent the breakdown of jobs with different position types or class levels. The second and third patterns visually encode the number of jobs in each career sector by circular spans and outer radius respectively. To ensure effective visualization, we ran a focus group with five career counselors from Rutgers Career Services. The counselors rejected the second design because it showed some bias on small career sectors. For the first design, some of them thought presenting every career sectors with equal size did not display the information visually and the bars outside the circle were confusing. Our final design adopts the third design pattern of career sectors.

7.3.4 User Interface and Interactions

JobViz is a web-based application implemented using D3.js (https://d3js.org/), a JavaScript library for data visualization. The user interface (as shown in Figure 7.1) consists of four components,

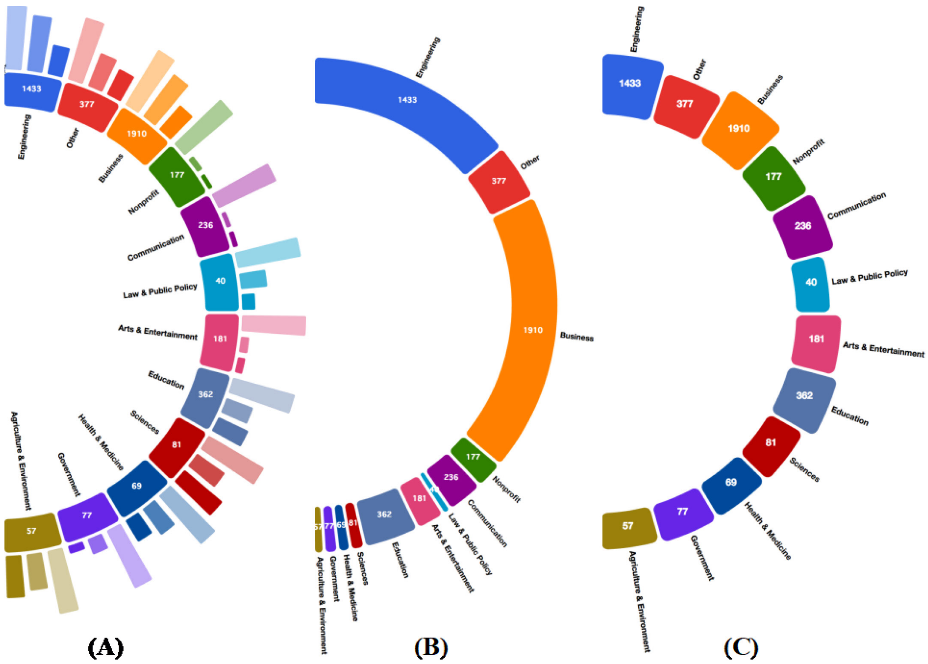


Figure 7.3: Three choices to design the career sectors. (a) was not adopted because of the visual clutter. (b) was rejected by the career counselors due to under representing small career sectors. (c) is the final design choice.

including a control panel and an inset of statistical view on the left module, a central module for data visualization display, and a right module for listing and filtering job postings. The user interface can support the following user interactions:

- **Load job posting data**. Users can select job posting data by date (e.g., 2015, 2016) from a dropdown menu and load the data for visualization.

- **Initiate/change visualization layout**. Users can group majors by school or by major area and color-code all of the major-career links by major groupings.

- **Query relationships**. Users can see links connect major(s) to related career sectors by selecting majors or a career sector in the visualization, or by searching for major(s) directly from a textbox on the control panel.

- Request tag cloud. Users can click the button titled "Tag Cloud" to display a tag cloud for a group of job postings and click the button "Tag Color" to color code the visualized tags by career sectors.

- Change the tension of links. Users can move a slider on the control panel to change the bundling effect on the links. The higher the value is, the closer the links will be.

- **Browse a Statistical View**. Users can switch between the tabs titled "Majors" and "Jobs" to see the statistics.

- Filter job postings. Users can filter the job postings listed on the right module by position type or class level.

- Open a career sector. Users can double click a career sector and see the industry sectors it contains.

- Open/Save a job posting. Users can identify a job from the visualization, open it to read the full posting, and save it in a "shopping cart" for later use.

7.3.5 System Overview

A flowchart of the system overview is shown in Figure 7.4. The data visualization contains multiple views that are well coordinated to accomplish the design principles (***DP1-4***): A Database View that provides an overview of all majors, job postings in the database, and their relationships (***DP1, 4***); a Relation View that visualizes the relationships between majors and jobs as spline curves connecting a set of majors with career sectors (***DP2***); and a Tag Cloud View that displays frequent keywords of job postings and links them to their sources visually using color (***DP3***). Details of the data can be accessed from the visualization (***DP2***). The multiple views of the visualization are visually encoded from three data processing steps triggered by user interactions in exploring the visualization: visualization layout creation, major-career links selection, and tag cloud computation. A statistical view is linked to the visualization to show the distributions of active majors and jobs.

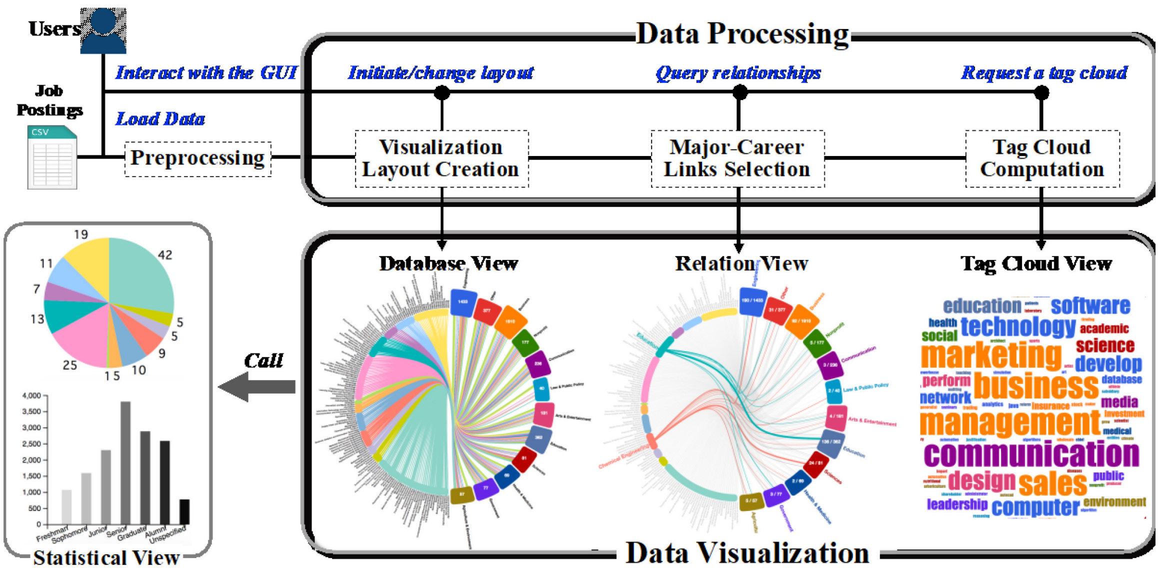


Figure 7.4: A system overview of JobViz. The system has a data processing part and a visualization part with three coordinated views: Database View, Relation View, and Tag Cloud View. Additional statistics of majors and jobs can be found in a Statistical View.

7.4 Algorithm Design

7.4.1 Preprocessing

The job posting data is currently downloaded from the career portal and uploaded as a CSV file into Jobviz. (Eventually the two systems could be integrated, but currently they are two separate systems). A linked list is then constructed for efficient set operations on the data. In the linked list, a job node stores all of the data fields (title, employer, position type, post time, etc.) and a reference to the requested majors. A major node contains the information of name, school, major area, and a reference to all of the relevant job nodes. Job descriptions of each job posting are parsed into single words (common words are removed) for future tag cloud computation and these words are added to each job node as a new field named "keywords".

7.4.2 Visualization Layout Creation

This algorithm computes the positions of majors, the size of career sectors, and the links between them in the visualization. It first clusters majors and arranges them alphabetically in each cluster (initially by school, but that can be changed to major area by users), then categorizes job nodes into industry sectors (or job categories), based on the information provided by the companies. These industry sectors are further grouped into twelve career sectors [CF]. As shown in Figure 7.3c, the outer radius of each career sector is computed as the following formula where the inner radius is constant for all career sectors:

$$OuterRadius_i = InnerRadius + 50 * log_2(number\ of\ jobs)$$

After the circular layout is plotted, the algorithm creates a contingency matrix L of which rows stand for majors and columns for career sectors. All of the job nodes are scanned and if a job from career sector n is requesting major m, the element l_{mn} in the matrix L will add a reference to this job. After the scanning, each element l_{mn} in the matrix will contain references to major m, career sector n, and a set of job nodes $\{J_{mn}\}$:

$$\{J_{mn}\} = \{jobs \in Career\ Field\ n\} \cap \{jobs\ requesting\ Major\ m\}$$

The visualization plots major-career links based on the matrix L. Whenever the set of jobs $\{J_{mn}\}$ in an element l_{mn} is not empty, a link will be created between major m and career sector n. These links are initially represented as gray spline curve with equal thickness, but can be colored to match the major groups with different thicknesses indicating the number of connected jobs.

7.4.3 Major-Career Links Selection

This algorithm will be triggered when users click a major or a career sector from the visualization or search for a major from the search text box on the left module. The algorithm first searches all of the major-career links to find those connected to the selected items, then identifies the corresponding elements in matrix *L* to extracts the job nodes from these elements. For each found link that connects major m to career sector n, the algorithm computes a correlation coefficient $Coef_{mn}$:

$$Coef_{mn} = \frac{|\{J_{mn}\}|}{|\{jobs \in Career\ Field\ n\}|}$$

The stronger the relationship between major *m* and career sector *n* is, the higher the $Coef_{mn}$ will be. In the visualization, all of the computed correlation coefficients are visually encoded as the thickness of the links whose colors follow those of selected majors or career sectors.

7.4.4 Tag Cloud Computation

This algorithm computes a tag cloud for the job nodes that are currently listed on the right module. It first collects all of the keywords from these job nodes and counts the frequency of each keyword, then sorts the keywords by frequencies. The first fifty (can be changed) keywords are selected to display as tags. Each tag links to the job postings that mention it. The position of each tag is randomly assigned and the tag size is computed from the frequency:

$$TagSize_i = \frac{InnerRadius}{8 * frequency_0} \times frequency_i$$

where the $frequency_0$ is the maximum frequency found for the set of job nodes.

7.5 User Scenarios

In this section, we illustrate JobViz's functionality and utility using two typical user scenarios with the key tasks that the tool is designed to support. The demonstration data are job postings posted from January to September in 2016.

7.5.1 Scenario One

In this scenario, student A is a freshman just starting to work with a career counselor and wants to understand the relationship between different majors and job opportunities. From the initial Database View (Figure 7.5a), she finds that the most popular major is actually "any major", i.e., finds out that many of the jobs being posted on the career portal do not really care about the major. Next, she looks at all of the major-career links by major areas as shown in Figure 7.5b. In this view,

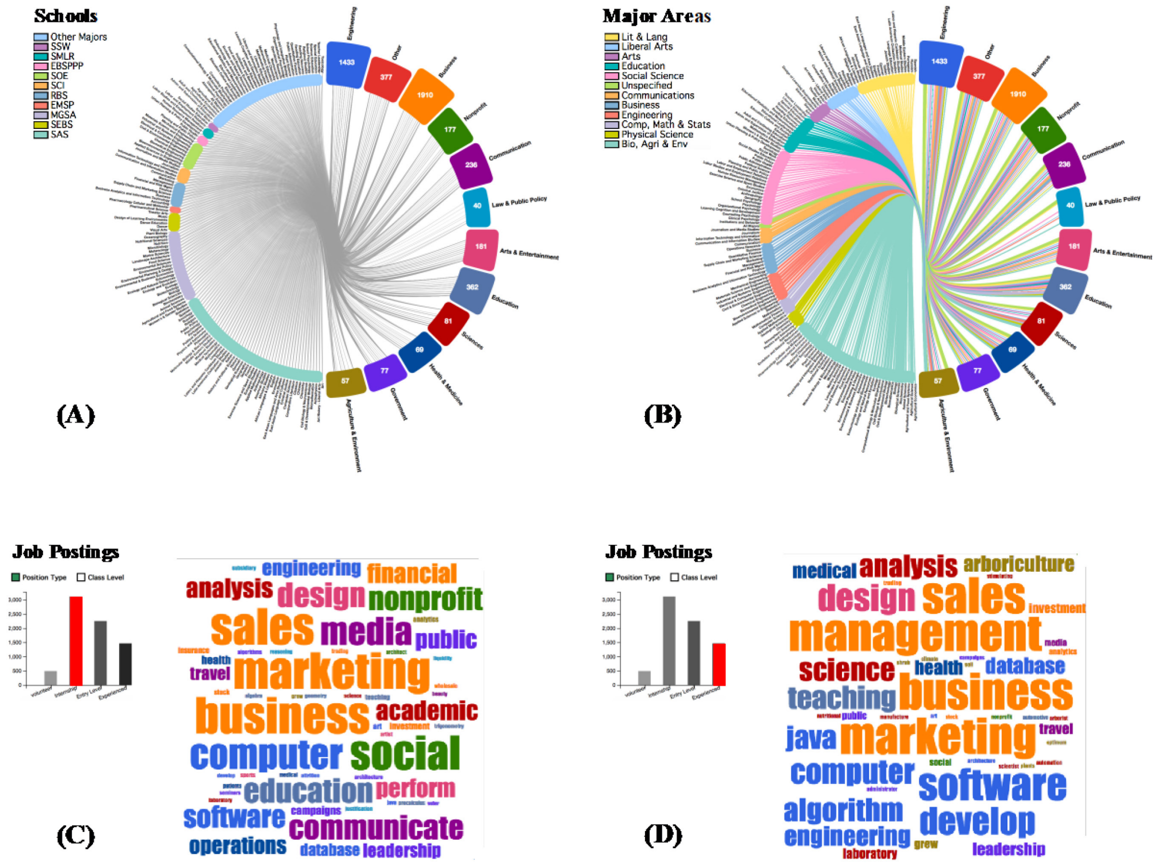


Figure 7.5: An overview of the job posting data posted in 2016. The initial Database View in (a) illustrates the distribution of jobs and major-career links. The Database View in (b) uses density and thickness of links to show the popularity of majors in the job market. The Tag Cloud Views in (c) and (d) summarize keywords of internships and full-time experienced jobs respectively.

several "job-popular" majors standout such as Education, Communication, Marketing, Computer Science, and Food Science. The career sectors "business" and "engineering" also show higher connectivity.

The student then explores the actual job postings that are listed on the right module. The histogram of job posting in the Statistical View indicates that more than half of the opportunities are internships (many job postings contain more than one position type). To further analyze, she filters the job postings by position type and requests tag clouds to show the most frequently mentioned key words in the internships and full-time jobs. From the two generated Tag Cloud Views (Figure 7.5c, d), she notices that they both contain the key words “business”, “sales”, “marketing”, and “computer”. However, some obvious distinctions are observed. The Tag Cloud View of internships contains some unique terms like “social”, “nonprofit”, “public”, and “education” which suggest many internships are in the nonprofit/education sector. On the other hand, “management”, “develop”, “algorithm”, “science”, “teaching”, and “java” are in the other Tag Cloud View, indicating that full-time experienced jobs are more related to majors and professional skills. (Note: JobViz contains only job postings from the university career portal, where employers are typically soliciting students with little or no experience.) Moreover, the color of tags indicates that “Engineering” and “Business” career sectors are more dominant in full-time experienced jobs.

7.5.2 Scenario Two

In this scenario, student B is a freshman or sophomore who wants to know the future employment in several career sectors and which majors could lead to those directions. This information can help him choose courses. He starts by clicking career sectors “Communication”, “Arts & Entertainment”, and “Science” successively to see three Relation Views and the associated Statistical Views (Figure 7.6). The visualizations indicate that a career sector relates to many more majors and even major areas than he expected, though each of them primarily requests majors from several targeted major areas (e.g., the career sector “Arts & Entertainment” welcomes social science majors while the “Science” sector mostly requests STEM majors). He opens the career sector “Arts & Entertainment”

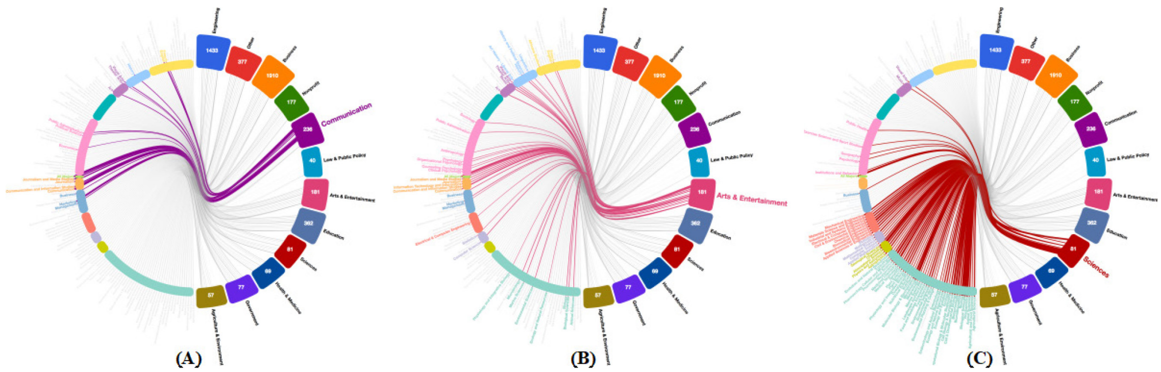


Figure 7.6: Three different Relation Views showing links from the chosen career sectors to majors listed in those job postings. (a) "Communication" sector and (b) "Arts & Entertainment" sector mainly focus on majors of arts, liberal arts, and social science, while (c) "Science" sector mostly recruits from STEM majors.

to see the industry sectors (job categories) it contains and then requests a Tag Cloud View to summarize the contained job postings (Figure 7.7). When he clicks the term "design" in the Tag Cloud View, a pop-up table lists all of the job postings that mention "design". After exploring the career sectors, he selects five majors to view their relationships to career sectors (Figure 7.1). The Relation View tells that majors (e.g., history) do not lock a student into a specific career sector (e.g., education) as he previously assumed.

7.6 User Study

To determine how well our design meets the design tasks and the usability of the system, we conducted a series of user studies. During the design phase, we consulted with potential users on visualization layout, demonstrated an initial prototype to additional potential users, and ran several focus groups on the second prototype (see Chapter 7.3). The effectiveness of the final design and

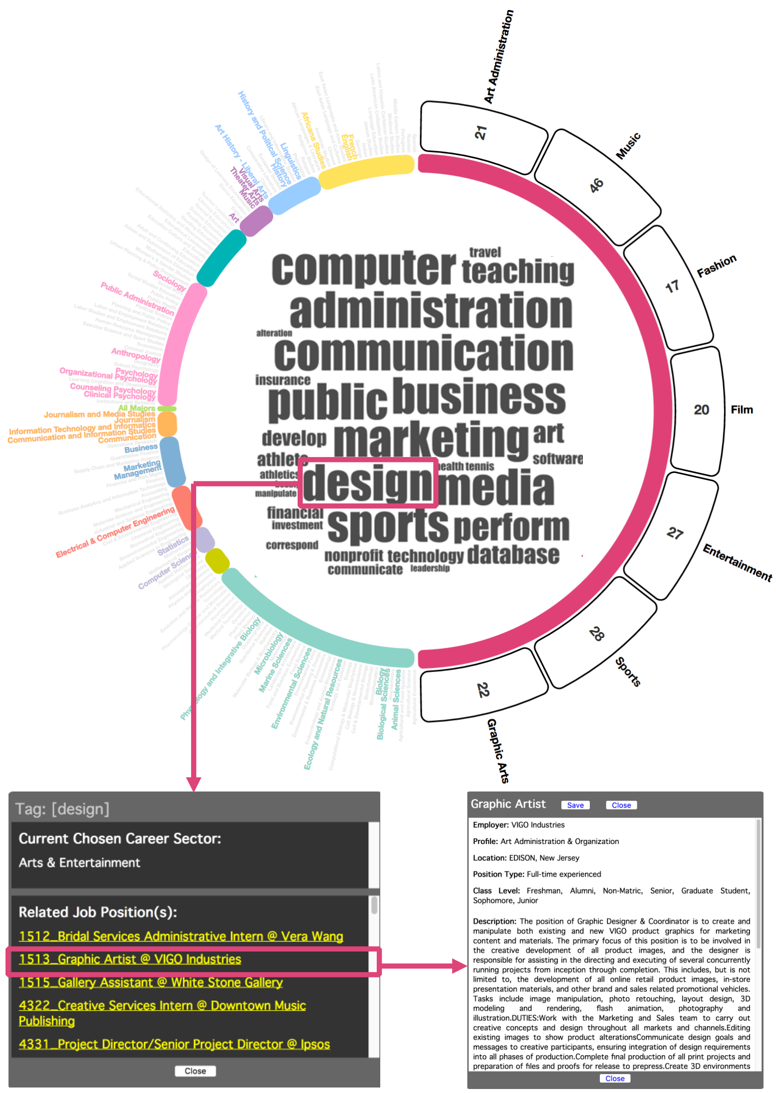


Figure 7.7: A Tag Cloud View for career sector "Arts & Entertainment". Each term in the tag cloud is linked to job postings.

the utility of the system was evaluated in two questionnaire-based studies: (a) 52 students who were recruited at Rutgers University and (b) 14 students from a career development course.

7.6.1 Questionnaire Details

Both studies used the same questionnaire which contains two sections. In the first section, participants were asked to rate JobViz based on ten criteria to evaluate the fulfillment of the design tasks (T1-5), novelty, interactivity, and usability. The seven criteria which relate to the design tasks are: (1) Providing an overview of the whole job database and its dynamics (T1); (2) Grouping majors and jobs (T1, 3); (3) Seeing intangible relationships between jobs and majors (T2); (4) Finding a job position of interest from the visualization (T3); (5) Providing details about individual job positions (T4); (6) Classifying jobs by different levels (T3); (7) Summarizing the features of majors and career fields (job categories) by tag cloud (T5). All ten criteria were rated on a modified Likhart-scale: 1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = excellent.

In the second section, participants were asked to answer four open-ended questions including "How does JobViz compare with the university career portal or other job search tools in helping you understand the job data?", "What did you learn about the relationships between jobs and majors?", "What impressions did you discern from the tag cloud?", and "What additional features would you like to see?"

7.6.2 Study Protocol

In the first study, we invited about 70 undergraduate students through asking in person, via email, or via social media, and 52 of them took part in our study within a three-month period. The demographic information of the student participants can be found in Table 7.1. Most of the students had previously used the university's career portal. In the classroom study, the JobViz tool was presented to a new career development course with 14 undergraduate students, in the context of general job search and career decision tools. The course aims to help students build the cognitive skills and reflective, exploratory approach that is essential for leading a rich inner and public life, as well as for building a satisfying career.

Table 7.1. Demographic Information of the Participants

	Recruited Students	**Students from the Class**
Gender	**24 males, 28 females**	**5 males, 9 females**
Age	20.4 ± 1.3	21.2 ± 1.1
Major	**Electrical Engineering, Computer Science, Accounting, Economics, Biology, Geology, Psychology, Mathematics, Physics, Arts, and Unspecified**	

The study consists of multiple sessions with different group sizes. Each study session was started with a short oral introduction of the study and consent form signing. After that, a four-minute DEMO video gave a brief demonstration on the JobViz and explained some important features. Then students were asked to complete three specific tasks (similar to the user scenarios in Chapter 7.5) to make sure they learn the tool sufficiently. Next, students used the user interface to explore the data based on their individual interests. Finally, students were invited to provide their written responses on the questionnaire. The classroom study put the study as an extra credit project and allowed student participants to submit their questionnaires electronically after class.

The submitted questionnaires are analyzed and reported in this way: (a) We analyze and compile the ratings to the ten criteria across the participants and report the statistics. (b) For the open-ended questions, we collate similar responses and report selected anonymous quotes or paraphrases to represent typical responses.

7.6.3 Results

Figure 7.8 summarizes all of the students' ratings of JobViz on the ten criteria (green-colored bars for the 52 recruited students and orange-colored bars for 14 students in the class). Statistically, two study groups gave similar ratings on eight out of ten criteria (P-value > 0.05).

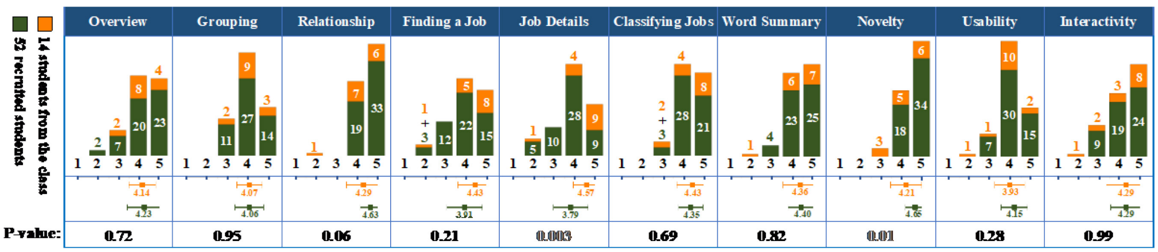


Figure 7.8: Results of questionnaire of both the recruited students and classroom study. The results of both studies are very similar.

Recruited Students

JobViz performance received ratings of over 4 (good) on eight out of ten items and over 3 (fair) on the remaining two (classifying jobs and finding a job) from the 52 recruited students. The two lower rated criteria may reflect the students' expressed desire for a more integrated tool with direct access to job postings. In the written responses to the "suggestion for additional features", we found that many students requested the ability to apply for jobs directly from the tool and some students wished to filter or sort jobs by job location and salary level. (Note: This first iteration of JobViz was not intended to be used to apply for a job.)

The written responses were generally positive (the compiled responses contain less than five negative statements) although only 36 students wrote responses to the open-ended questions. 32 out of these 36 responses expressed a positive attitude towards JobViz over other job search tools in helping understand the job data. More than half (22) of the written responses indicated that the participants had changed or improved their understanding of the job market. A representative response is: "*I learned that there is a basic skill set that is needed for the job, majoring in science does not necessarily mean I will be using it for the jobs, which is also shown by the tag cloud.*"

Classroom Study

Similarly, the students in the classroom study rated JobViz's performance highly (>4) on nine out of ten criteria with a better than fair rating in usability.

11 out of 14 students preferred JobViz in responding to the first open-ended question. Frequently mentioned strengths of JobViz include: comprehensive, interactive, efficient, visual, and overview. In a characteristic description, one student wrote "*One of the overall advantages of the JobViz is the tangible representation of job opportunities in correlation to a varied amount of majors. ... Students aren't pigeon-holed by their major.*" Negative comments referred to the lack of detailed information about jobs, odd connections to particular job postings, the lack of updates in the posting database, and the learning curve in using JobViz.

In addition, 13 students expressed that using JobViz changed their understanding of the relationship between jobs and majors. A representative comment notes "*your major does not determine what your job will be.*" Many students pointed out that majors relate to more different jobs than they previously realized: "*I saw less of a correlation between jobs and majors than I would have previously expected. I learned that for any one major, there are numerous possible careers. Likewise, for any one career, there are countless possible majors.*" Another student wrote

"*When I click "economics", I received a plethora of results other than the generic business jobs I'm used to seeing on most other sites.*"

Students associated the tag cloud with job-related skills: "*It overtly shows what areas of specialization are most prevalent. This can demonstrate valuable associated skills in the area.*" A negative comment was raised saying "*Several words such as network, communicate, public, and resources were very general.*"

7.6.4 Discussion of User Study

The results of both user studies confirm the fulfillment of the established design tasks (**T1-7**) and support the utility of JobViz. In both studies, the average rating was "good" with little statistical difference between criteria or groups (Figure 7.8). Comments tended towards positive attitudes. Students were generally impressed with the visual connections and dynamic interactivity in JobViz.

The small size of the study groups (compared to the student population) and the reliance on expressed opinions limit the interpretation of our results. A more detailed and longitudinal study would be needed to track whether using a visualization tool like JobViz resulted in changes of majors and more qualitative discussions between counselors, professors, and students.

However, the study suggests that students can benefit from JobViz to gain an understanding of the job market, of the relationships between majors and jobs, and of the skills desired in potential employees independent of majors. The students themselves expressed realization that many of their prior assumptions were incorrect. In particular, a common prior assumption was that a major prepares for a particular career path; in contrast, JobViz shows that a given major leads to a variety of career paths (Figure 7.1). Similarly, students previously assumed that an industry sector usually requires job seekers from a specific major area, but in JobViz, they see that most jobs request a range of majors and some even indicate that any major is acceptable (Figure 7.6). These insights are almost impossible to glean from the standard querying interface. Career counselors are aware of them but expressed difficulty communicating such insights. JobViz thus fulfills its primary goal as a teaching tool.

One caution is the students' enthusiasm to take the links simplistically (one student comment mentioned "dream job" and there were other similar if less flamboyantly stated comments), which suggests a need for ongoing guidance from career counselors. However, we observed no tendency for students to shift majors towards engineering, despite its high connection to jobs. In fact, the students seemed comforted to know there are many opportunities but they need to be aware of the skills and requirements. Given the initial positive feedback, JobViz is being incorporated as a standard component of the career development courses.

7.7 Lessons Learned

Although we designed JobViz specifically for visualizing job postings, the insights gained from the development of a teaching tool and the lessons learned from the project are valuable for the visualization community and researchers in a range of application domains.

Multivariate categorical data can be visualized by many different set visualization techniques, and our choices of techniques were driven by the need to display connections between five variables (majors, job titles, position types, class levels, and job descriptions) and by the cardinality of those variables (100’s to 1000’s of majors and job titles). Our first design (Figure 7.2) following existing approaches [ID90, KSB09, KBH06] resulted in a busy figure with little emphasis on the major-to-job connections critical to our users’ needs. Instead, we combined an aggregation approach (grouping job titles into career sectors) with a circular layout of Contingency Wheel style visualizations, bundling the connection lines to focus visual energy on the connections.

Reduction of bias motivated a number of design choices. The circular layout deemphasizes the importance of position for both majors and career sectors relative to alternative linear layouts. The equal span of career sectors is intended to promote focus on all career sectors. The initial presence of the connection curves for every major-career link reduces the effects of selection bias by focusing attention first on the broader patterns; the continued presence of connection curves in faded grey reminds users of these broader patterns. Testing for reduction in bias would be of great value, and some approaches could include a large design study exposing different groups to different designs or a long-term longitudinal study tracking behavior based on timing of exposure to a bias-reducing visualization. However, these tests are beyond the scope of our current work and could be part of a future endeavor.

Our interactions with career counselors in the Rutgers Career Services facilitated our understanding of the data and the user requirements from their perspective of many years of counseling students. This helped us improve our designs with a teaching and education-centric focus. For example, in our second prototype we adopted the counselors’ suggestion to group jobs into twelve career sectors representing equal-spanning circular sectors (see the discussion in Chapter 7.3). This project exemplifies a case where the tension between designing a tool for intuitive use by novice users unfamiliar with the data (students) and ensuring the satisfaction of experts (career counselors) using or suggesting the tool as a teaching device influenced the visualization design choices.

As stated previously, many students told us that they realized that their assumptions about the relationships between jobs and majors turned out to be wrong. Even though the students had previous experience using popular job search engines in addition to the university career service portal, they perceived a different story from JobViz. For them, looking at the "big picture" formed a positive viewpoint-changing experience. Two factors contribute to this result: (a) the presentation of an overview designed to simultaneously satisfy the critical viewpoint of experts (here the career counselors) and yet also depict an intuitive "big picture" view of the data to novices (here the students) and (b) a relative neutral visual design that deemphasizes positioning and selection bias.

The overall system has several limitations in its current implementation. Firstly, it is not directly connected to the university career portal. The data is imported into the system. This limitation will be addressed in the next implementation. Secondly, the categorization of job postings relies on the industry sectors (employers) to self-categorize. This is a general problem for the university systems as the employers are not very accurate about their own characterizations. Thirdly, the current tag cloud computation parses phrases (e.g., machine learning) into single words. This will also be fixed in future releases. In the future, we plan to deploy JobViz as a GUI to the university career portal and make the tool available to the whole Rutgers community.

7.8 Conclusion

In today's data rich environment, everyone theoretically has access to data, but access alone does not make the data easy to understand. Most jobs in the US are listed online, but students who are novices in the field of job data analysis need assistance to extract the information for informed decisions about their majors, minors, and future careers. An expert summary or job article fails to satisfy the demand for individualized information. JobViz fills this gap by providing a visual overview of the full breadth of career sectors and complete map of job-major connections while still allowing user interaction and filtering. The key lesson here is that overview images are difficult to create but are immensely important for communicating the data to novice users. Querying only works when one knows what to ask. Queries can also lead to biased views of the data as only the information requested is displayed. Well-designed overviews present a holistic and big picture view of the data, creating truly accessible information.

Acknowledgement

We wish to thank Rutgers Career Services for providing us job posting data as well as many useful feedbacks along the way. We also thank all students who participated in user testing and the final questionnaire studies.

Chapter 8: Discussions

8.1 General Usages of the Pipeline Model in Case Studies

The creation of overview visualizations in the two case studies (Chapter 6 and 7) implicitly follows the proposed pipeline model in Chapter 4. The rest of this section discusses how each case study fits into the pipeline model and the characterization of overviews. The next section will discuss specific observations leading out of the overall work of this thesis.

8.1.1 Case Study One

a. Visual Design Pipeline

A flowchart of the visual design pipeline is shown in Figure 8.1. The pipeline starts from compiling a bunch of system requirements through structured interviews and working with ocean scientists who are the primary targeted users of the system-to-be. These requirements reflect scientists' concern and the way a visualization can help them improve their understanding of ocean circulation systems. Based on the system requirements, we set a visualization goal.

At the second level (identifying design principles and the basics of overview), we first identified four design principles (DP1 – DP4 as can be found in Figure 8.1) to cover the system requirements. Then the nature of overviews (the first dimension of the characterization, see Figure 3.1 and Chapter 3.1) in the visualization-to-be was decided to convey the shapes, distribution, properties, and motions of eddies (overview of content and overview of change) according to the first two system requirements. Finally, the roles of overview (the second dimension of the characterization, R1 – R5 in Chapter 3.2) were derived from the design principles: 1. The second

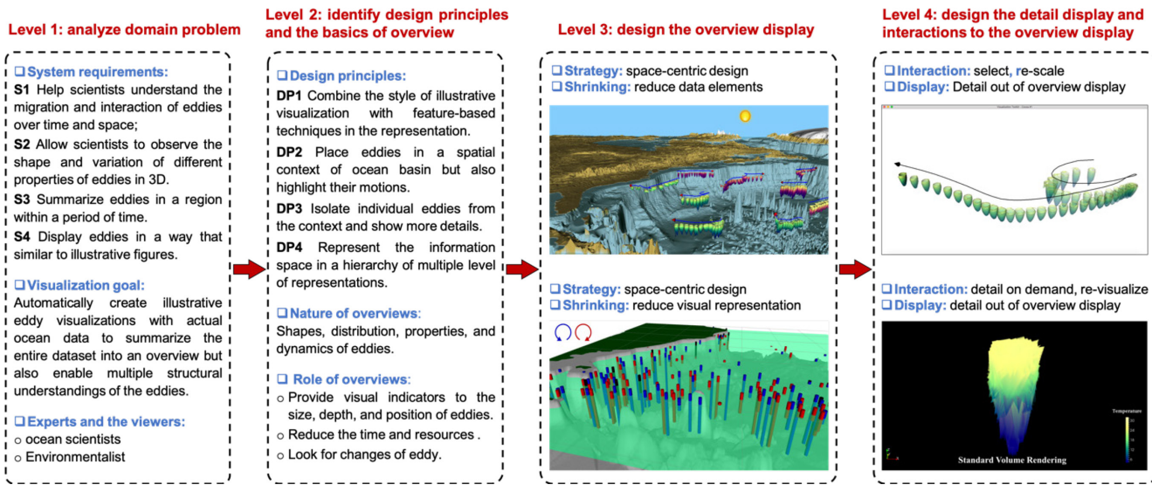


Figure 8.1: The visual design pipeline of the first case study: eddy visualization.

design principle (DP2) requires the overviews to play a role in monitoring or looking for the changes of eddies (R5); 2. DP3 requires the overviews to provide visual indicators to the size, depth, and locations of a large number of eddies (R2); 3. DP4 requires the overviews to reduce the time and resources needed to navigate the information space (R3).

At the third level (designing the overview displays), we adopted a space-centric design strategy (the third dimension of the characterization, see Figure 3.1 and Chapter 3.3) which uses a 3D ocean basin as context and overlays eddies on top of the context. This design strategy complies with the second design principle (DP2). Considering the characteristics of the ocean simulation data and the nature of overviews identified in the previous level, two types of overview display were designed to match the basics of overviews identified at the second level: 3D Eddy Path Map and Eddy Skeleton View. The 3D Eddy Path Map highlights the size, shape, and motion of eddies while the Eddy Skeleton View constructs a concise overview of the orientation, depth, and distribution of eddies. Two shrinking approaches (the fourth dimension of the characterization, see

Figure 3.1 and Chapter 3.4) were applied to each of the overview display: a. extracting high-vorticity regions from raw data as eddy instances and filter the visualized eddies by their durations (reduce the number of data elements); b. representing eddy instances as cylinders to highlight their locations, orientations, and depths (reduce visual representation).

In designing the detail display and interactions to the overview display (the fifth dimension of the characterization, see Figure 3.1 and Chapter 3.5) at the fourth level, three major interaction schemes to the overview displays were designed: (1) filter the presented eddy instances from the Eddy Skeleton View by a specified property (e.g., the durations of eddy as shown in Figure 6.7); (2) isolate an individual eddy path from Eddy Path Map and re-scale the eddy path in an Isolated Eddy Path View (as shown in Figure 6.8 and 6.9); (3) select an individual eddy instance from an Isolated Eddy Path View, re-visualize the eddy in an Individual Eddy View (as shown in Figure 6.10), and allow more details on demand. The Isolated Eddy Path View and Individual Eddy View are the two detail displays of the overview visualization of eddies.

b. Data Flow Pipeline

A flowchart of the data flow pipeline is shown in Figure 8.2. The data pipeline starts from studying the essential information of the ocean simulation data such as file format (NetCDF file), duration (50 years), sampling rate (one data frame per day), variables, and the coordinate system.

In data characterization at the second level, the ocean simulation was characterized as time-varying volume data based on the data study and domain analysis at the first level. The Okubo-Weiss (OW) criterion was chosen to isolate eddies from the ocean simulation data. A series of key

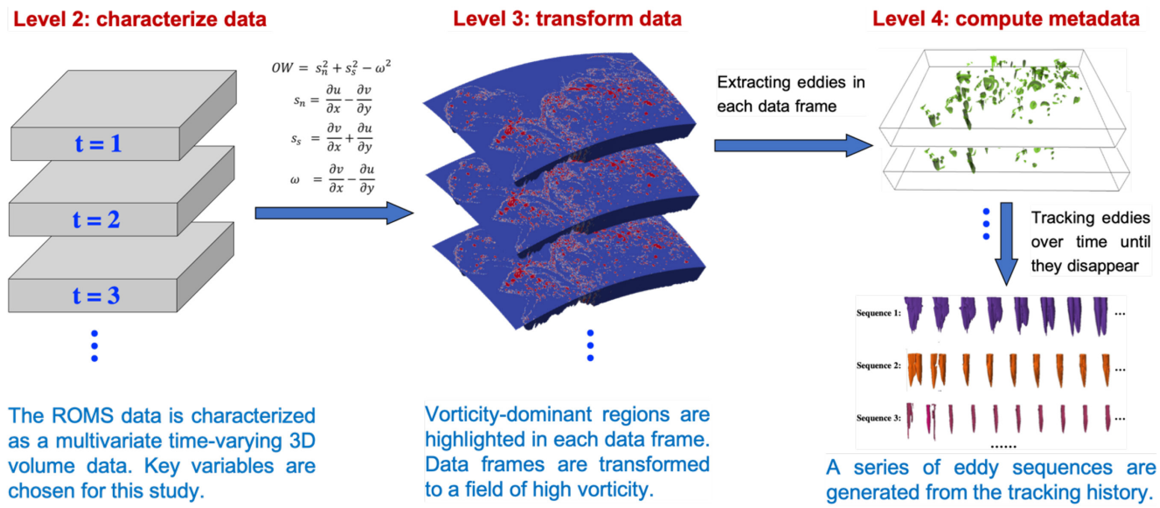


Figure 8.2: The data flow pipeline of the first case study: eddy visualization.

variables were identified to compute the OW parameters and physical properties (e.g., temperature, salinity, velocity, PH) of eddies in this study.

At the third level (data transformation), we compute OW parameter and the values of specified physical properties at each coordinate and thus transformed each data frame to a vector field. A field of the OW parameter represents the distribution of vorticity at different regions.

At the fourth level (metadata computation), vorticity-dominant regions were extracted from each data frame by applying a threshold to the OW field of the data frame. These extracted regions are ocean eddies. Several geometrical and physical properties were computed for each of the extracted ocean eddy at every time steps. These eddies were tracked over time by a feature tracking algorithm. A series of eddy paths were then queried and created from the tracking history and the metadata of eddy extraction. As a summary, the general usage of the pipeline model in the first case study is shown in Figure 8.3.

Level 1:	✓ **Research Problem**: The structure of eddies is 3D yet most research is still focused on 2D plots of eddy variables on the surface and in vertical section. ✓ **Potential Viewers/Users**: Oceanographers, environmentalists, and meteorologists ✓ **System Requirements**: S1 – S4. ✓ **Visualization Goal:** Automatically produce illustrative 3D visualizations of ocean eddies with actual data. ✓ **Data**: A numerical simulation of the northwest Atlantic Ocean off the coast of North America generated by the Regional Ocean Modeling System (ROMS). The dataset is saved as the NetCDF format and the sampling rate is one data frame per day. The variables, dimensions, and coordinate system are studied.	
Level 2:	✓ **Nature of Overviews**: content and change. ✓ **Roles of Overviews:** R2, R3, R5. ✓ **Design Principles**: DP1 – DP4.	✓ The ROMS data are characterized as multivariate time-varying volume data. ✓ The evolution of ocean eddies can be modeled as turbulence in vector fields over time.
Level 3:	✓ **Design Strategies of Overview Displays**: Use 3D ocean basin as context and overlay eddy evolutions on top of the background (space-centric design). ✓ **Shrinking Approaches**: reduce data element (3D Eddy Path Map); reduce visual representation (Eddy Skeleton View).	✓ The Okubo-Weiss (OW) criteria is chosen to divides the ocean into vorticity dominant regions and the background. ✓ The raw data is transformed into a vector fields (contains multiple variables).
Level 4:	✓ **Interaction Paradigm**: Overview and detail on demand with multiple separate views and navigational slaving coordination. ✓ **Detail Display Design**: visual emphasis by color and transparency; separate display with identical representation.	✓ Compute OW parameter and the other specified variable at each coordinate. ✓ Extract eddies from each data frame. ✓ Track eddies over time. ✓ Query and create eddy paths from the tracking history and the metadata of extraction.
Level 5:	✓ **Platform**: Linux, MacOS ✓ **Application**: A GUI implemented in Qt and VTK and written with standard C++ language. ✓ **Functionality**: Four illustrative visualizations are provided to visualize the computed eddy metadata at different levels from multiple eddy paths in a 3D map with ocean basin to a single eddy feature.	

Figure 8.3: The general usage of the pipeline model in the first case study.

8.1.2 Case Study Two

a. Visual Design Pipeline

A flowchart of the visual design pipeline is shown in Figure 8.4. The design pipeline starts from a requirement analysis. The targeted users of the system-to-be include college students, professors, and career counselors. As part of our design process, we worked closely with the career counselors at Rutgers Career Services and interviewed with professors and students to collect their requirements. Generally, potential users would like to see the relationships between college majors and job opportunities and a concise summary of job postings. Based on the requirement analysis, a

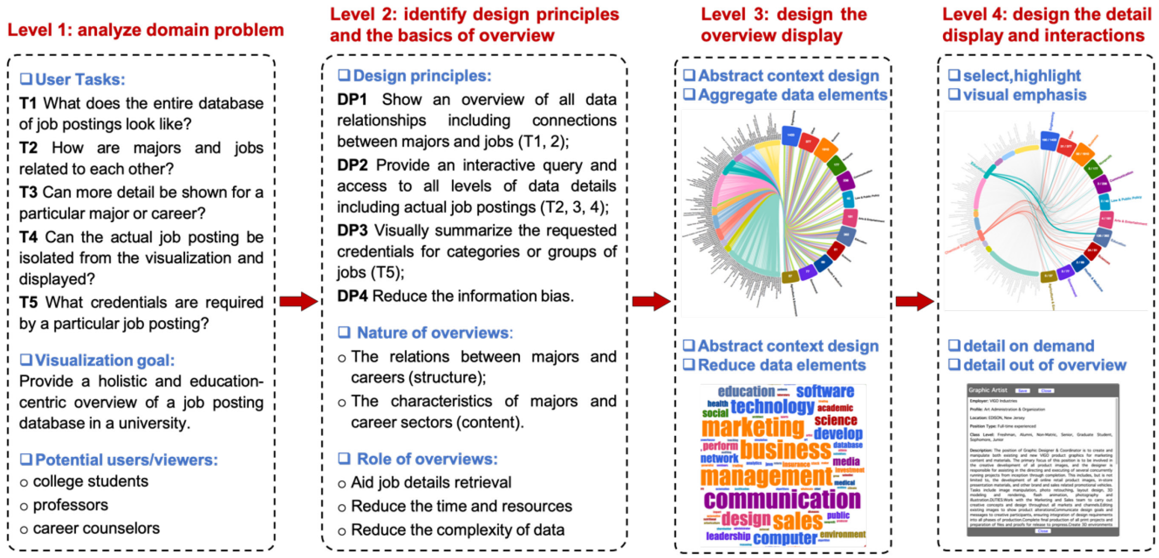


Figure 8.4: The visual design pipeline of the second case study (job visualization).

visualization goal was set as "providing a holistic and education-centric overview of a job database in a university". We further compiled a list of user analysis tasks (see Chapter 7.3) which summarize the topics, concerns, and sentiments expressed in the interviews.

At the second level (identifying design principles and the basics of overview), we first identified four design principles (DP1 – DP4) to confirm the user analysis tasks in the following design process. Then the nature of overviews (see Chapter 3.1 and Figure 3.1) in the visualization-to-be was decided to convey the relations between majors and careers (an overview of structure) and the characteristics of majors and job groups (an overview of content). Finally, the roles of overview (R1 – R5 in Chapter 3.2) were derived from the design principles: 1. DP1 requires the overviews to reduce the complexity of data and show intangible relationships between majors and jobs (R4); 2. DP2 requires the overviews to provide visual indicators to assist job details retrieval (R1); 3. DP3 requires the overviews to summarize a collection of jobs to reduce the time of navigation (R3).

At the third level (designing the overview displays), we adopted an abstract context design strategy. Considering the characteristics of the data and the nature of overviews identified in the previous level, we designed two overview displays: Relation Overview and Tag Cloud View. A Relation Overview combines the intuitiveness of node-link diagrams and the scalability of aggregation-based techniques to construct a summary of the relationships between majors and career sectors. The Tag Cloud View uses a text representation to summarize a group of job postings by key words in different sizes and colors. Two shrinking approaches were applied to each of the overview display: (1) categorizing job postings into different career sectors (aggregating data elements); (2) visualizing the most frequent mentioned terms from job descriptions (reduce the number of data elements).

At the fourth level (designing the detail display and interactions to the overview display), two major interaction schemes to the overview displays were designed: (1) select a major (or a career sector) to highlight its connections to career sectors (or majors); (2) isolate a job posting from the visualization to get more information about the job. These two interactions generate two detail displays of the overview visualizations: Single Relation View and Job Posting View.

b. Data Flow Pipeline

A flowchart of the data flow pipeline is shown in Figure 8.5. The data pipeline starts from studying the job posting data from Rutgers University Career Service. The database stores the data entries in a tabular format and often contains about 10, 000 job postings.

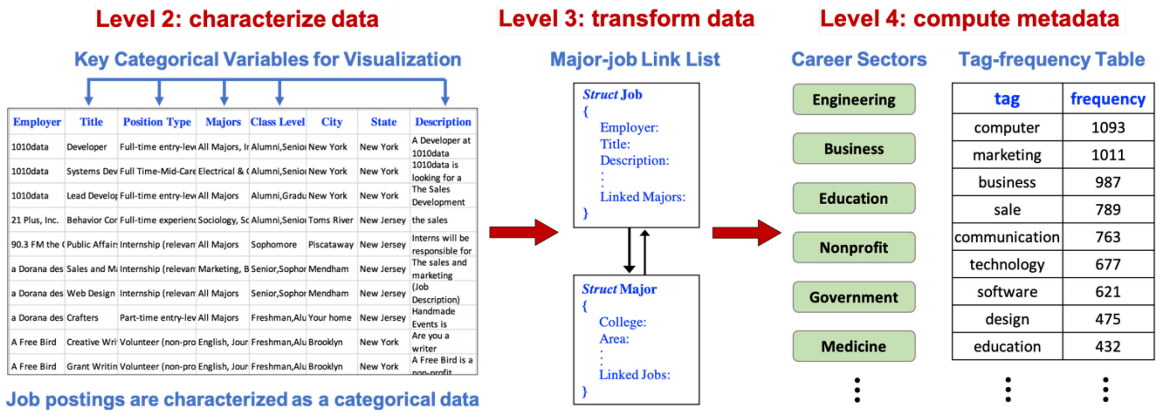


Figure 8.5: The data flow pipeline of the second case study (job visualization).

In data characterization at the second level, the job posting data was characterized as a categorical data that contains a series of variables whose values comprise a set of discrete categories. The domain problem is equivalent to displaying large numbers of categories and data entries and the intangible relationships between categorical variables. Five key variables (i.e., Job Title, Position Type, Majors, Class Level, Job Descriptions) were chosen for visualization.

At the third level (data transformation), the job posting data was transformed to a major-job linked list for efficient set operations. In the linked list, a job node stores all of the data fields (title, employer, position type, post time, etc.) and a reference to the requested majors. A major node contains the information of name, school, major area, and a reference to all of the relevant job nodes.

At the fourth level (metadata computation), two metadata (analytical abstraction) were computed for visualization: career sectors and tag-frequency table. Driven by the third design principle (DP3), job postings were categorized into twelve career sectors. Each of these career sectors is represented as a circular sector with a unique color in the visualization. Tag-frequency table is used to plot a Tag Cloud View. It contains a column of keywords extracted from job descriptions and a column of frequencies of the keywords. The table was sorted by frequencies. As a summary, the general usage of the pipeline model in the first case study is shown in Figure 8.6.

Level 1:	✓ **Research Problem**: Current job search and assessment methods can't assist students to understand the links between college majors and job opportunities and provide effective career counseling. ✓ **Potential Viewers/Users**: college students, university career counselors, and faculty. ✓ **User Tasks:** T1 – T5. ✓ **Visualization Goal**: Providing a holistic and education-centric overview of a job database. ✓ **Data**: Job postings from the university career service database. The database stores the data entries in a tabular format and often contains about 10, 000 job postings. The essential properties are studied.	
Level 2:	✓ **Nature of Overviews**: the relations between majors and careers (structure); the characteristics of majors and career sectors (content). ✓ **Roles of Overviews:** R1, R3, R4. ✓ **Design Principles**: DP1 – DP4.	✓ Job postings are characterized as multivariate categorical data. ✓ The domain problem is modeled as displaying categories and data entries and the relationships between categorical variables.
Level 3:	✓ **Design Strategy**: node-link diagram, contingency wheel, and tag cloud (abstract context design). ✓ **Shrinking Approaches**: job posting entries are aggregated into career sectors which are represented as circular sectors in the visual design (aggregate data elements).	✓ Four categorical variables are chosen to study the interconnection between them. ✓ Data entries are grouped into categories. ✓ The raw job posting data is reduced into a bunch of interconnected sets.
Level 4:	✓ **Interaction Paradigm**: Focus in the context ✓ **Detailed Visualization Design**: visual emphasis by color and transparency (major-job links); space distortion (fisheye effect).	✓ Classify jobs into career sectors; ✓ Create a multiple linked-list of majors and jobs; ✓ Extract keywords from job titles and job descriptions.
Level 5:	✓ **Platform**: any OS ✓ **Application**: A web-based visualization application is implemented using D3.js, a JavaScript library for data visualization. ✓ **Functionality**: The data visualization contains multiple views that are well coordinated to accomplish the design principles: a Database View, a Relation View, and a Tag Cloud View. Details of the data can be accessed from the visualization.	

Figure 8.6: The general usage of the pipeline model in the second case study.

8.2 Observations from Case Studies

Based on the two case studies in this thesis, the following observations can be made about overviews in data visualization:

(1) Overviews often lead to an improved satisfaction from the visualization and a better understanding of data. There are at least four reasons that contribute to this improvement. First, overviews aid viewers in extracting the gist of data more accurately and rapidly. For example, finding which college majors are more job-popular from JobViz is much easier than searching and querying a job list. Second, overviews are by nature general summaries or descriptions and often put the objects of interest in a context (e.g., the 3D Eddy Path Map in Figure 6.5). Third, overviews provide users with directly manipulable control mechanisms that facilitate exploration and allow users to focus attention on the larger information problem at hand. Lastly, viewers tend to make decisions without perfect information. People seldom take the time to act optimally, often "satisficing" rather than optimizing [Sim81]. In the second case study, many students told us that they realized that their assumptions about the relationships between jobs and majors turned out to be wrong. Even though the students had previous experience using popular job search engines in addition to the university career service portal, they perceived a different story from JobViz. For them, looking at the "big picture" formed a positive viewpoint-changing experience. Figure 8.7 gives a comparison of displaying job postings as a long list in the Rutgers Career Portal and showing an overview visualization of job postings in the JobViz system.

(2) An effective overview needs to balance the abstractness and exhaustiveness. On one hand, an overview is a display that shrinks an information space and shows information about it at a coarse level of granularity [HH11]. On the other hand, overviews should be fairly exhaustive to represent the most salient features of data. A bad example of balancing abstractness and exhaustiveness in overview visualization is the initial prototype in the development process of

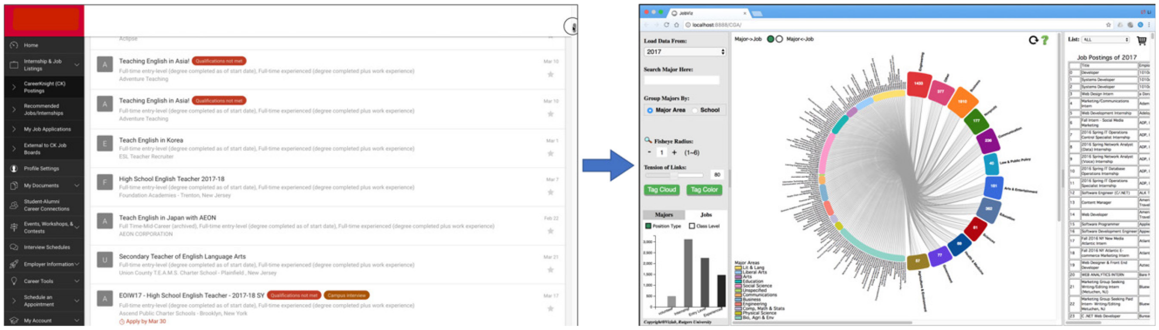


Figure 8.7: A comparison of a list of job postings (left) and an overview of job postings (right).

JobViz (the second case study). As shown in Figure 8.8, the visualization displays a large number of college majors, job titles, and the relations between majors and jobs and thus become very cluttered. The overview is over exhaustive. A good example of balancing the abstractness and

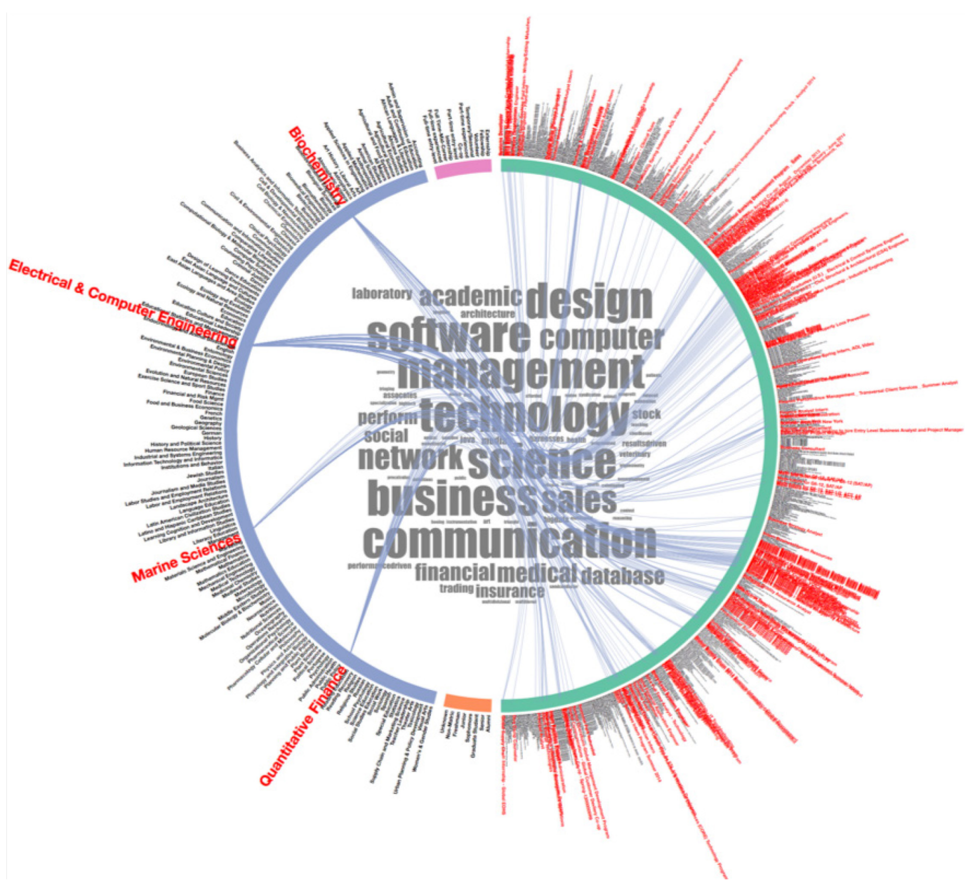


Figure 8.8: The initial prototype of JobViz exemplifies an over exhaustive overview.

exhaustiveness is the Eddy Skeleton View. As shown in Figure 8.9, the visualization gives a concise visual summary of the distributions and heights of cyclonic and anticyclonic eddies in a region.

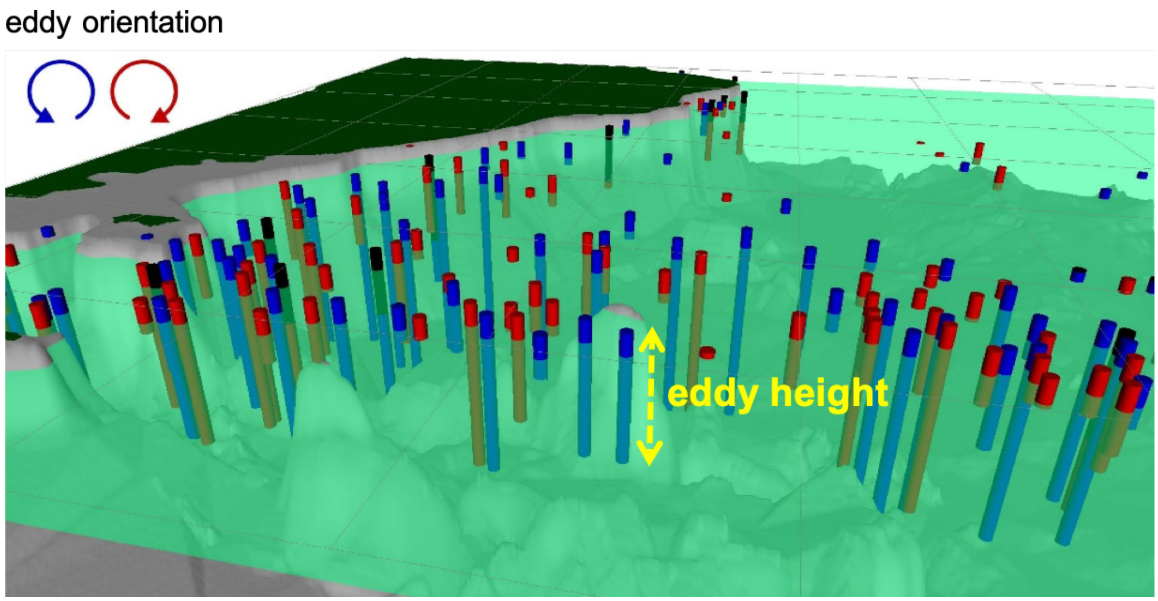


Figure 8.9: The Eddy Skeleton View (the first case study) exemplifies a good example of balancing the abstractness and exhaustiveness in an overview.

(3) The design of an overview display is both application and data dependent. The design strategy of an overview display is usually accordance with the data type in scientific visualization while designers have more control over the design strategy of an overview display in information visualization. The roles of overviews are dominantly affected by the application domains. Relatively speaking, shrinking approaches are mostly data dependent in both scientific visualization and information visualization. The design process of the JobViz (the second case study) exemplifies this observation (Figure 8.10). Considering the characteristics of categorical data (i.e., a categorical variable comprises a set of discrete categories), majors were clustered by school or major area and jobs were categorized into twelve career sectors. The design decisions of

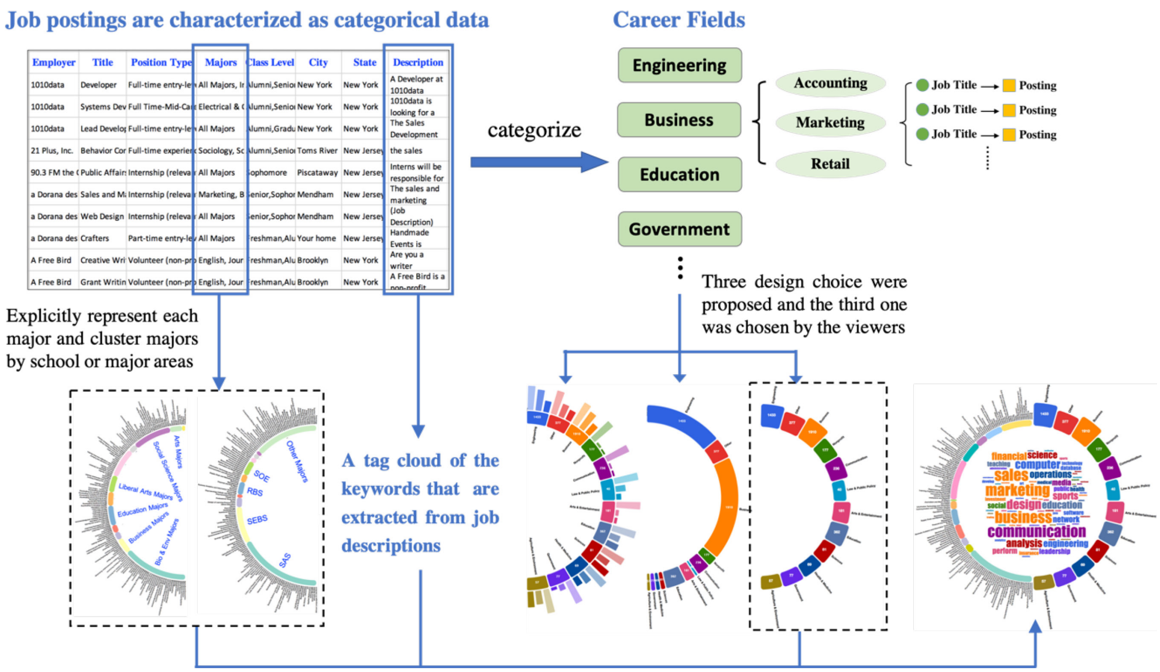


Figure 8.10: The design process of the JobViz (the second case study) shows that design choices depend on both application domain (e.g., users' desire) and the characteristics of data.

explicitly representing and arranging majors along with a half-circle, summarizing jobs with a tag cloud in the center of the circular layout, and representing career sectors as equal-spanning circular sectors with unique color were primarily decided by the favors or requirements of users.

(4) Context can strengthen or even change the awareness of overviews. As shown in Figure 8.11, the ocean basin and sea shore in the 3D Eddy Path Map construct a context to highlight the orientations, positions, sizes, and motions of eddies and thus strengthen the awareness of "change" from the overview. In Figure 8.12, the presence of connection curves and major names in faded grey constructs a context of the highlighted visual components in the visualization. The context reminds viewers of a broader patterns of major-career connections and gives users a reference to how many majors and connections are highlighted compare with those in total.

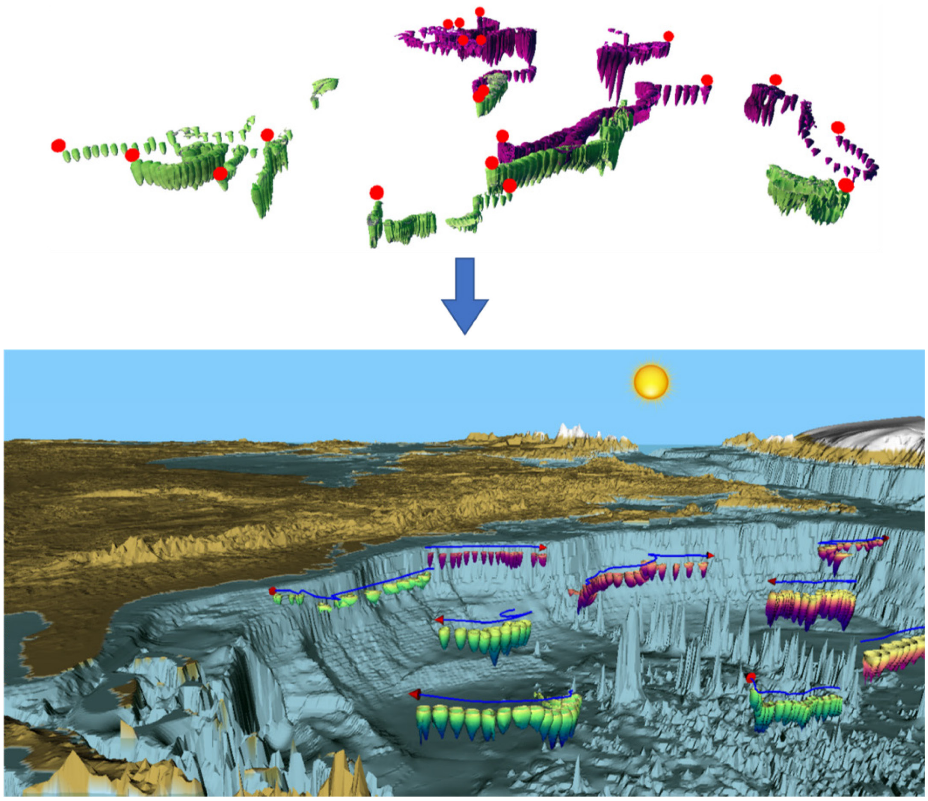


Figure 8.11: The ocean basin and sea shore in the 3D Eddy Path Map construct a context to highlight the orientations, positions, sizes, and motions of eddies and thus strengthen the awareness of "change" from the overview.

(5) Overviews can play an important role in bias reduction and helping viewers form an anchoring to the data. Anchoring is the phenomenon that a previous stimulus provides a frame of reference [VZS18]. As stated in Chapter 7.7, the neutral visual design in the overviews deemphasizes positioning and selection bias and thus increases the exposure of previously under-representative college majors and jobs. The "holistic and neutralized overviews of the data" change students' previous assumptions about the relationships between jobs and majors and give the students an anchoring of estimation to the data which can aid their future judgement.

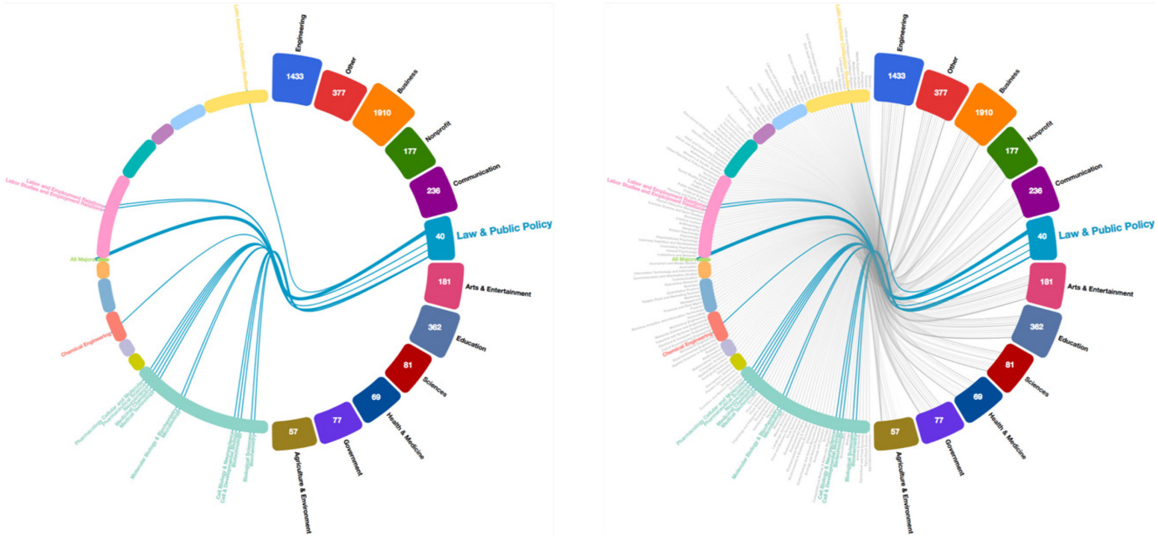


Figure 8.12: The grey major names and major-career links in the JobViz construct a context to give references and reminders to viewers.

These observations may be specific to the applications, tasks, and users concerned. Nevertheless, they do offer takeaway messages for visualization designers to consider in creating overview visualizations in different domains. The results and evaluations of the two case studies in this thesis validate that overview visualizations can facilitate data understanding and analysis in both scientific visualization and information visualization.

Chapter 9: Conclusion and Future Work

This thesis focuses on "overview" visualizations as a means to covey the context of a large dataset. An overview visualization is a visual representation that provides viewers with an overall awareness of the content, structure, or changes of the data (the dataset can contain time-varying information) while allowing the viewer to further drill down into the details. The applicability of overview visualizations is extended to both scientific visualization and information visualization domains. Overview visualizations are characterized into five dimensions to help visualization designers better understand the five most important aspects of overview visualization. Based on the characterization, a pipeline model is proposed to guide the development of overview visualizations or to analyze the existing visualizations. Two case studies with evaluations in scientific visualization and information visualization are used to demonstrate the integrated design process of overview visualizations and how overview visualizations can give viewers a better understanding on the data and an improved satisfaction from the visualizations.

In future work, I wish to carry out a formal user study to validate the characterization of overview visualization in Chapter 3 and the pipeline model in Chapter 4. Moreover, the relation between overview and detail needs further work. More knowledge is needed about how different overview designs are useful for different kinds of task, about the relative contributions of the global and local features of a visual scene in creating an overview, and about the role of interaction in creating the awareness of overviews. Finally, I plan to expand the application of the proposed characterization and pipeline model to more case studies.

Chapter 10: Summary of Accomplishment and Related Publications

The content of this thesis has already been published or is to be published in a number of papers and posters. The first case study was published in an IEEE SciVis poster [4] and the Computer Graphics Forum journal [5]. The activity detection algorithm to isolate eddy paths from the results of eddy extraction and tracking was published in a paper of IEEE Symposium on Large Data Analysis and Visualization (LDAV) [1]. The second case study has been published in an IEEE InfoVis poster [3] and in the IEEE Computer Graphics and Application magazine [6]. The contents in Chapter 2 and Chapter 3 are going to be submitted to a EuroVis State-of-the-Art Report [7] as a survey paper.

My other research work outside the scope of this thesis include extracting skeletons from 3D volume data, visualizing hydrothermal plume in the ocean [2], and the visualization of time-varying scientific data in web browsers [8].

Journal Publication:

[1] Bemis, Karen G., Deborah Silver, Guangyu Xu, Russ Light, Darrell Jackson, Christopher Jones, Sedat Ozer, and Li Liu. "The path to COVIS: A review of acoustic imaging of hydrothermal flow regimes." Deep Sea Research Part II: Topical Studies in Oceanography 121 (2015): 159-176.

[2] Liu, Li, Deborah Silver, Karen Bemis, Dujuan Kang, and Enrique Curchitser. "Illustrative Visualization of Mesoscale Ocean Eddies." In Computer Graphics Forum, vol. 36, no. 3, pp. 447-458. 2017.

[3] Liu, Li, Deborah Silver, and Karen Bemis. "Application-Driven Design: Help Students Understand Employment and See the “Big Picture”." IEEE computer graphics and applications 38, no. 3 (2018): 90-105.

Conference Publication:

[4] Liu, Li, Sedat Ozer, Karen Bemis, Jay Takle, and Deborah Silver. "An interactive method for activity detection visualization." In Large-Scale Data Analysis and Visualization (LDAV), 2013 IEEE Symposium on, pp. 127-130. IEEE, 2013.

[5] Liu, Li, Deborah Silver, and Karen Bemis, "JobViz: Interactive Visualization of Majors & Jobs", Poster Session, IEEE InfoVis 2016, October 2016.

[6] Bemis, Karen, Li Liu, Deborah Silver, Dujuan Kang, Enrique Curchitser, "Case Study on Visualizing Gulf Stream Eddies from ROMS", Poster Session, IEEE SciVis 2016, Best Poster Winner, October 2016.

In Preparation:

[7] Liu, Li, Deborah Silver, and Karen Bemis. “Overviews in Data Visualization.” (will be submitted to EuroVis State-of-the-Art Reports)

[8] Liu, Li, Deborah Silver, and Karen Bemis. “Interactive Visualization of Time-Varying Scientific Data in Web Browsers” In IEEE Computer Graphics and Applications. (in the process)

Reference

[AAD10] Andrienko, Gennady, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. "Space, time and visual analytics." International Journal of Geographical Information Science 24, no. 10 (2010): 1577-1600.

[AAMG12] Alsallakh, Bilal, Wolfgang Aigner, Silvia Miksch, and M. Eduard Gröller. "Reinventing the contingency wheel: Scalable visual analytics of large categorical data." *IEEE Transactions on Visualization and Computer Graphics* 18, no. 12 (2012): 2849-2858.

[AWB17] Abdul-Rahman, Alfie, Simon Walton, Karen Bemis, Julie Gonnering Lein, Katharine Coles, Deborah Silver, and Min Chen. "Re-spatialization of time series plots." Information Visualization 16, no. 4 (2017): 275-290.

[ABF07] Agrafiotis, Dimitris K., Deepak Bandyopadhyay, and Michael Farnum. "Radial clustergrams: visualizing the aggregate properties of hierarchical clusters." Journal of chemical information and modeling 47, no. 1 (2007): 69-75.

[AES05] Amar, Robert, James Eagan, and John Stasko. "Low-level components of analytic activity in information visualization." In Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on, pp. 111-117. IEEE, 2005.

[AF06] Appert, Caroline, and Jean-Daniel Fekete. "OrthoZoom scroller: 1D multi-scale navigation." In Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 21-30. ACM, 2006.

[AMA14] Alsallakh, Bilal, Luana Micallef, Wolfgang Aigner, Helwig Hauser, Silvia Miksch, and Peter Rodgers. "Visualizing sets and set-typed data: State-of-the-art and future challenges." In *Eurographics conference on Visualization (EuroVis)–State of The Art Reports*, pp. 1-21. 2014.

[AS04] Amar, Robert, and John Stasko. "A knowledge task-based framework for design and evaluation of information visualizations." In IEEE Symposium on Information Visualization, vol. 2004, pp. 143-49. Citeseer, 2004.

[BBB11] Boyandin, Ilya, Enrico Bertini, Peter Bak, and Denis Lalanne. "Flowstrates: An approach for visual exploration of temporal origin-destination data." In Computer Graphics Forum, vol. 30, no. 3, pp. 971-980. Oxford, UK: Blackwell Publishing Ltd, 2011.

[BC13] Burigat, Stefano, and Luca Chittaro. "On the effectiveness of Overview+Detail visualization on mobile devices." Personal and ubiquitous computing 17, no. 2 (2013): 371-385.

[BBS08] Botchen, Ralf P., Sven Bachthaler, Fabian Schick, Min Chen, Greg Mori, Daniel Weiskopf, and Thomas Ertl. "Action-based multifield video visualization." IEEE Transactions on Visualization and Computer Graphics 14, no. 4 (2008): 885-899.

[BCP12] Brambilla, Andrea, Robert Carnecky, Robert Peikert, Ivan Viola, and Helwig Hauser. "Illustrative flow visualization: State of the art, trends and challenges." Visibility-oriented Visualization Design for Flow Illustration (2012).

[Ber18] Desiderata for Knowledge Graphs, http://www.mkbergman.com/, February 21, 2018.

[BGT] Burning Glass Technologies, http://burning-glass.com/.

[Big13] Bigml, Using Text Analysis To Predict H-1B Wages: http://blog.bigml.com/2013/10/01/using-text-analysis-topredict-h1-bwages/, Oct. 2013.

[BSP93] Bier, Eric A., Maureen C. Stone, Ken Pier, William Buxton, and Tony D. DeRose. "Toolglass and magic lenses: the see-through interface." In Proceedings of the 20th annual conference on Computer graphics and interactive techniques, pp. 73-80. ACM, 1993.

[BSX15] Bemis, Karen G., Deborah Silver, Guangyu Xu, Russ Light, Darrell Jackson, Christopher Jones, Sedat Ozer, and Li Liu. "The path to COVIS: A review of acoustic imaging of hydrothermal flow regimes." Deep Sea Research Part II: Topical Studies in Oceanography 121 (2015): 159-176.

[BW90] Beard, David V., and JOHN Q. WALKER II. "Navigational techniques to improve the display of large two-dimensional spaces." Behaviour & Information Technology 9, no. 6 (1990): 451-466.

[BWF10] Born, Silvia, Alexander Wiebel, Jan Friedrich, Gerik Scheuermann, and Dirk Bartz. "Illustrative stream surfaces." *IEEE transactions on visualization and computer graphics* 16, no. 6 (2010): 1329-1338.

[BWP10] Bremer, Peer-Timo, Gunther Weber, Valerio Pascucci, Marc Day, and John Bell. "Analyzing and tracking burning structures in lean premixed hydrogen flames." *IEEE Transactions on Visualization and Computer Graphics* 16, no. 2 (2010): 248-260.

[CCF97] Carpendale, M. Sheelagh T., David J. Cowperthwaite, and F. David Fracchia. "Extending distortion viewing from 2D to 3D." IEEE Computer Graphics and Applications 17, no. 4 (1997): 42-51.

[CF] Career Fields, https://career.berkeley.edu/InfoLab/CareerFields.

[CJR07] Caban, Jesus, Alark Joshi, and Penny Rheingans. "Texture-based feature tracking for effective time-varying data visualization." IEEE Transactions on Visualization and Computer Graphics 13, no. 6 (2007): 1472-1479.

[CKB08] Cockburn, Andy, Amy K. Karlson, and Benjamin B. Bederson. "A review of overview+ detail, zooming, and focus+context interfaces." ACM Comput. Surv. 41, no. 1 (2008): 2-1.

[CLTE11] Chaigneau, Alexis, Marie Le Texier, Gérard Eldin, Carmen Grados, and Oscar Pizarro. "Vertical structure of mesoscale eddies in the eastern South Pacific Ocean: A composite analysis from altimetry and Argo profiling floats." *Journal of Geophysical Research: Oceans* 116, no. C11 (2011).

[CM97] Card, Stuart K., and Jock Mackinlay. "The structure of the information visualization design space." In Information Visualization, 1997. Proceedings., IEEE Symposium on, pp. 92-99. IEEE, 1997.

[CMN13] Clyne, John, Pablo Mininni, and Alan Norton. "Physically-Based Feature Tracking for CFD Data." *IEEE transactions on visualization and computer graphics* 19, no. 6 (2013): 1020-1033.

[CMS94] Cignoni, Paolo, Claudio Montani, and Roberto Scopigno. "MagicSphere: an insight tool for 3D data visualization." In Computer Graphics Forum, vol. 13, no. 3, pp. 317-328. Blackwell Science Ltd, 1994.

[CMS99] Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman, eds. Readings in information visualization: using vision to think. Morgan Kaufmann, 1999.

[CR98] Chi, Ed Huai-hsin, and John T. Riedl. "An operator interaction framework for visualization systems." In Information Visualization, 1998. Proceedings. IEEE Symposium on, pp. 63-70. IEEE, 1998.

[CSC06] Correa, Carlos, Deborah Silver, and Min Chen. "Feature aligned volume manipulation for illustration and visualization." IEEE Transactions on Visualization and Computer Graphics 12, no. 5 (2006).

[DBSL07] Doglioli, A. M., Bruno Blanke, Sabrina Speich, and Guillaume Lapeyre. "Tracking coherent structures in a regional ocean model with wavelet analysis: Application to Cape Basin eddies." *Journal of Geophysical Research: Oceans* 112, no. C5 (2007).

[DC03] Daniel, Gareth, and Min Chen. "Video visualization." In Proceedings of the 14th IEEE Visualization 2003 (VIS'03), p. 54. IEEE Computer Society, 2003.

[DGH03] Doleisch, Helmut, Martin Gasser, and Helwig Hauser. "Interactive feature specification for focus+ context visualization of complex simulation data." In VisSym, vol. 3, pp. 239-248. 2003.

[DH02] Doleisch, Helmut, and Helwig Hauser. "Smooth brushing for focus+ context visualization of simulkation data in 3D." (2002).

[Dol07] Doleisch, Helmut. "SimVis: Interactive visual analysis of large and time-dependent 3D simulation data." In Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come, pp. 712-720. IEEE Press, 2007.

[Dro11] Drocourt, Yoann, Rita Borgo, Kilian Scharrer, Tavi Murray, S. I. Bevan, and Min Chen. "Temporal visualization of boundary-based geo-information using radial projection." In Computer Graphics Forum, vol. 30, no. 3, pp. 981-990. Oxford, UK: Blackwell Publishing Ltd, 2011.

[DTW15] Duffy, Brian, Jeyarajan Thiyagalingam, Simon Walton, David J. Smith, Anne Trefethen, Jackson C. Kirkman-Brown, Eamonn A. Gaffney, and Min Chen. "Glyph-based video visualization for semen analysis." IEEE transactions on visualization and computer graphics 21, no. 8 (2015): 980-993.

[EDG08] Elmqvist, Niklas, Thanh-Nghi Do, Howard Goodell, Nathalie Henry, and Jean-Daniel Fekete. "ZAME: Interactive large-scale graph visualization." In Visualization Symposium, 2008. PacificVIS'08. IEEE Pacific, pp. 215-222. IEEE, 2008.

[Emsi] Emsi, http://www.economicmodeling.com/services/.

[Figshare13] Figureshare, Geometric zooming versus semantic zooming. https://figshare.com/articles/_Geometric_zooming_versus_semantic_zooming_/650434, March 13, 2013.

[FS00] Friendly, Michael, and SAS Institute. *Visualizing categorical data*. Cary, NC: SAS Institute, 2000.

[GABJ08] Gosink, Luke, John C. Anderson, E. Wes Bethel, and Kenneth I. Joy. "Query-driven visualization of time-varying adaptive mesh refinement data." IEEE transactions on visualization and computer graphics 14, no. 6 (2008): 1715-1722.

[Gap12] Gapminder, The Gapminder World Map, August 2012, http://www.gapminder.org/.

[GEO02] Grant, Josh, Gordon Erlebacher, and James O'Brien. "Case study: visualizing ocean flow vertical motions using Lagrangian-Eulerian time surfaces." In *Proceedings of the conference on Visualization'02*, pp. 529-532. IEEE Computer Society, 2002.

[GMPS00] Greene, Stephan, Gary Marchionini, Catherine Plaisant, and Ben Shneiderman. "Previews and overviews in digital libraries: Designing surrogates to support visual information seeking." Journal of the American Society for Information Science 51, no. 4 (2000): 380-393.

[GNBP11] Gasteiger, Rocco, Mathias Neugebauer, Oliver Beuing, and Bernhard Preim. "The FLOWLENS: A focus-and-context visualization approach for exploration of blood flow in cerebral aneurysms." IEEE Transactions on Visualization and Computer Graphics 17, no. 12 (2011): 2183-2192.

[GTG11] Garcia, Juan, Roberto Theron, and Francisco Garcia. "Semantic zoom: A details on demand visualization technique for modelling owl ontologies." In Highlights in Practical Applications of Agents and Multiagent Systems, pp. 85-92. Springer, Berlin, Heidelberg, 2011.

[GWG03] Gross, Tom, Wido Wirsam, and Wolfgang Graether. "AwarenessMaps: visualizing awareness in shared workspaces." In CHI'03 Extended Abstracts on Human Factors in Computing Systems, pp. 784-785. ACM, 2003.

[Hau06] Hauser, Helwig. "Generalizing focus+ context visualization." In *Scientific visualization: The visual extraction of knowledge from data*, pp. 305-327. Springer, Berlin, Heidelberg, 2006.

[HBP01] Hornbæk, Kasper, Benjamin B. Bederson, and Catherine Plaisant. "Navigation patterns and usability of overview+ detail and zoomable user interfaces for maps." proc. of the ACM Transactions on Computer-Human Interaction 9, no. 4 (2001): 362-389.

[HCL05] Heer, Jeffrey, Stuart K. Card, and James A. Landay. "Prefuse: a toolkit for interactive information visualization." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 421-430. ACM, 2005.

[HF03] Hornbæk, Kasper, and Erik Frøkjær. "Reading patterns and usability in visualizations of electronic documents." ACM Transactions on Computer-Human Interaction (TOCHI) 10, no. 2 (2003): 119-149.

[HH11] Hornbæk, Kasper, and Morten Hertzum. "The notion of overview in information visualization." International Journal of Human-Computer Studies 69, no. 7-8 (2011): 509-525.

[HHB15] Ha, Hyoji, Wonjoo Hwang, Sungyun Bae, Hanmin Choi, Hyunwoo Han, Gi-Nam Kim, and Kyungwon Lee. "CosMovis: Semantic Network Visualization by Using Sentiment Words of Movie Review Data." In Information Visualisation (iV), 2015 19th International Conference on, pp. 436-443. IEEE, 2015.

[HHN00] Havre, Susan, Beth Hetzler, and Lucy Nowell. "ThemeRiver: Visualizing theme changes over time." In Information visualization, 2000. InfoVis 2000. IEEE symposium on, pp. 115-123. IEEE, 2000.

[HMBG00] Hauser, Helwig, Lukas Mroz, Gian-Italo Bischi, and M. Eduard Gröller. "Two-level volume rendering—fusing MIP and DVR." In Proceedings of the conference on Visualization'00, pp. 211-218. IEEE Computer Society Press, 2000.

[HMBG01] Hauser, Helwig, Lukas Mroz, G. Italo Bischi, and M. Eduard Groller. "Two-level volume rendering." IEEE transactions on visualization and computer graphics 7, no. 3 (2001): 242-252.

[HMCM09] Hsu, Wei-Hsien, Jianqiang Mei, Carlos D. Correa, and Kwan-Liu Ma. "Depicting Time Evolving Flow with Illustrative Visualization Techniques." In *ArtsIT*, pp. 136-147. 2009.

[HMM00] Herman, Ivan, Guy Melançon, and M. Scott Marshall. "Graph visualization and navigation in information visualization: A survey." IEEE Transactions on visualization and computer graphics 6, no. 1 (2000): 24-43.

[Hol06] Holten, Danny. "Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data." *IEEE Transactions on visualization and computer graphics* 12, no. 5 (2006): 741-748.

[HR08] Hearst, Marti A., and Daniela Rosner. "Tag clouds: Data analysis tool or social signaller?." In *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, pp. 160-160. IEEE, 2008.

[ID90] Inselberg, A., and Bernard Dimsdale. "Parallel coordinates: a tool for visualizing multi-dimensional geometry.(1990)." *DOI: http://dx. doi. org/10.1109/VISUAL* (1990).

[Indeed] Indeed Ltd., http://www.indeed.com.

[JAM] Revealed impacts of smaller-scale eddies in northwest pacific on ocean circulation. http://www.jamstec.go.jp/e/about/press_release/20141215/

[JCRS09] Joshi, Alark, Jesus Caban, Penny Rheingans, and Lynn Sparling. "Case study on visualizing hurricanes using illustration-inspired techniques." IEEE transactions on visualization and computer graphics 15, no. 5 (2009): 709-718.

[JKM01] Jankun-Kelly, T. J., and Kwan-Liu Ma. "A study of transfer function generation for time-varying volume data." In Volume Graphics 2001, pp. 51-65. Springer, Vienna, 2001.

[JR05] Joshi, Alark, and Penny Rheingans. "Illustration-inspired techniques for visualizing time-varying data." In Visualization, 2005. VIS 05. IEEE, pp. 679-686. IEEE, 2005.

[JR08] Joshi, Alark, and Penny Rheingans. "Evaluation of illustration-inspired techniques for time-varying data visualization." In Computer Graphics Forum, vol. 27, no. 3, pp. 999-1006. Blackwell Publishing Ltd, 2008.

[JS06] Ji, Guangfeng, and Han-Wei Shen. "Feature tracking using earth mover's distance and global optimization." In Pacific Graphics, vol. 2. 2006.

[JSW03] Ji, Guangfeng, Han-Wei Shen, and Rephael Wenger. "Volume tracking using higher dimensional isosurfacing." In Proceedings of the 14th IEEE Visualization 2003 (VIS'03), p. 28. IEEE Computer Society, 2003.

[KSB09] Krzywinski, Martin, Jacqueline Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. "Circos: an information aesthetic for comparative genomics." *Genome research* 19, no. 9 (2009): 1639-1645.

[KBH06] Kosara, Robert, Fabian Bendix, and Helwig Hauser. "Parallel sets: Interactive exploration and visual analysis of categorical data." *IEEE transactions on visualization and computer graphics* 12, no. 4 (2006): 558-568.

[KC13] Kang, Dujuan, and Enrique N. Curchitser. "Gulf Stream eddy characteristics in a high-resolution ocean model." *Journal of Geophysical Research: Oceans* 118, no. 9 (2013): 4474-4487.

[Kin10] Kincaid, Robert. "Signallens: Focus+ context applied to electronic time series." IEEE Transactions on Visualization and Computer Graphics 16, no. 6 (2010): 900-907.

[KL06] Kincaid, Robert, and Heidi Lam. "Line graph explorer: scalable display of line graphs using focus+ context." In Proceedings of the working conference on Advanced visual interfaces, pp. 404-411. ACM, 2006.

[KMH01] Kosara, Robert, Silvia Miksch, and Helwig Hauser. "Semantic depth of field." In infovis, vol. 1, p. 97. 2001.

[KMH02] Kosara, Robert, Silvia Miksch, and Helwig Hauser. "Focus+ context taken literally." IEEE Computer Graphics and Applications 22, no. 1 (2002): 22-29.

[LA94] Leung, Ying K., and Mark D. Apperley. "A review and taxonomy of distortion-oriented presentation techniques." ACM Transactions on Computer-Human Interaction (TOCHI) 1, no. 2 (1994): 126-160.

[LEA92] Lancaster, Jack L., David Eberly, Abdalmajeid Alyassin, J. Hunter Downs, and Peter T. Fox. "A geometric model for measurement of surface distance, surface area, and volume from tomographic images." *Medical physics* 19, no. 2 (1992): 419-431.

[LPP06] Lee, Bongshin, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. "Task taxonomy for graph visualization." In Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization, pp. 1-5. ACM, 2006.

[LI95] Lokuge, Ishantha, and Suguru Ishizaki. "Geospace: An interactive visualization system for exploring complex information spaces." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 409-414. ACM Press/Addison-Wesley Publishing Co., 1995.

[LMC02] Lum, Eric B., Kwan-Liu Ma, and John Clyne. "A hardware-assisted scalable solution for interactive volume rendering of time-varying data." *IEEE Transactions on Visualization and Computer Graphics* 8, no. 3 (2002): 286-301.

[LR96] Lamping, Jonh, and Ramana Rao. "The hyperbolic browser: A focus+ context technique for visualizing large hierarchies." Journal of Visual Languages & Computing 7, no. 1 (1996): 33-55.

[LS08] Lu, Aidong, and Han-Wei Shen. "Interactive storyboard for overall time-varying data visualization." In *Visualization Symposium, 2008. PacificVIS'08. IEEE Pacific*, pp. 143-150. IEEE, 2008.

[LSB17] Liu, Li, Deborah Silver, Karen Bemis, Dujuan Kang, and Enrique Curchitser. "Illustrative Visualization of Mesoscale Ocean Eddies." In Computer Graphics Forum, vol. 36, no. 3, pp. 447-458. 2017.

[LSB18] Liu, Li, Deborah Silver, and Karen Bemis. "Application-Driven Design: Help Students Understand Employment and See the “Big Picture”." IEEE computer graphics and applications 38, no. 3 (2018): 90-105.

[LWL17] Liu, Dongyu, Di Weng, Yuhong Li, Jie Bao, Yu Zheng, Huamin Qu, and Yingcai Wu. "Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations." IEEE transactions on visualization and computer graphics 23, no. 1 (2017): 1-10.

[MAIS16] Matsuoka, Daisuke, Fumiaki Araki, Yumi Inoue, and Hideharu Sasaki. "A New Approach to Ocean Eddy Detection, Tracking, and Event Visualization–Application to the Northwest Pacific Ocean." *Procedia Computer Science* 80 (2016): 1601-1611.

[MH01] Mroz, Lukas, and Helwig Hauser. "RTVR-a flexible java library for interactive volume rendering." In Visualization, 2001. VIS'01. Proceedings, pp. 279-564. IEEE, 2001.

[MKO08] Muigg, Philipp, Johannes Kehrer, Steffen Oeltze, Harald Piringer, Helmut Doleisch, Bernhard Preim, and Helwig Hauser. "A Four-level Focus+ Context Approach to Interactive Visual Analysis of Temporal Features in Large Scientific Data." In Computer Graphics Forum, vol. 27, no. 3, pp. 775-782. Blackwell Publishing Ltd, 2008.

[MLF12] Ma, Kwan-Liu, Isaac Liao, Jennifer Frazier, Helwig Hauser, and Helen-Nicole Kostis. "Scientific storytelling using visualization." *IEEE Computer Graphics and Applications* 32, no. 1 (2012): 12-19.

[MM09] Muelder, Chris, and Kwan-Liu Ma. "Interactive feature extraction and tracking by utilizing region coherency." In *Visualization Symposium, 2009. PacificVis' 09. IEEE Pacific*, pp. 17-24. IEEE, 2009.

[Monster] Monster Ltd., http://monster.com/.

[MS00] Ma, Kwan-Liu, and Han-Wei Shen. "Compression and accelerated rendering of time-varying volume data." In *Proceedings of the 2000 International Computer Symposium-Workshop on Computer graphics and virtual reality*, pp. 82-89. 2000.

[Mun98] Munzner, Tamara. "Drawing large graphs with h3viewer and site manager." In Graph Drawing, pp. 384-393. Springer Berlin/Heidelberg, 1998.

[Mun09] Munzner, Tamara. "A nested model for visualization design and validation." *IEEE transactions on visualization and computer graphics* 15, no. 6 (2009).

[PBA10] Pietriga, Emmanuel, Olivier Bau, and Caroline Appert. "Representation-independent in-place magnification with sigma lenses." IEEE Transactions on Visualization and Computer Graphics 16, no. 3 (2010): 455-467.

[NOA] National oceanic and atmospheric administration. http://www.noaa.gov/

[Oku70] Okubo, Akira. "Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences." In *Deep sea research and oceanographic abstracts*, vol. 17, no. 3, pp. 445-454. Elsevier, 1970.

[OSBM14] Ozer, Sedat, Deborah Silver, Karen Bemis, and Pino Martin. "Activity detection in scientific visualization." *IEEE transactions on visualization and computer graphics* 20, no. 3 (2014): 377-390.

[OWS12] Ozer, Sedat, Jishang Wei, Deborah Silver, Kwan-Liu Ma, and Pino Martin. "Group dynamics in scientific visualization." In *Large Data Analysis and Visualization (LDAV), 2012 IEEE Symposium on*, pp. 97-104. IEEE, 2012.

[PCS95] Plaisant, Catherine, David Carr, and Ben Shneiderman. "Image-browser taxonomy and guidelines for designers." Ieee Software 12, no. 2 (1995): 21-32.

[PLB01] Pfister, Hanspeter, Bill Lorensen, Chandrajit Bajaj, Gordon Kindlmann, Will Schroeder, Lisa Sobierajski Avila, K. M. Raghu, Raghu Machiraju, and Jinho Lee. "The transfer function bake-off." *IEEE Computer Graphics and Applications* 21, no. 3 (2001): 16-22.

[PAK08] Purchase, Helen C., Natalia Andrienko, T. J. Jankun-Kelly, and Matthew Ward. "Theoretical foundations of information visualization." In Information Visualization, pp. 46-64. Springer, Berlin, Heidelberg, 2008.

[PWM13] Petersen, Mark R., Sean J. Williams, Mathew E. Maltrud, Matthew W. Hecht, and Bernd Hamann. "A three-dimensional eddy census of a high-resolution global ocean simulation." *Journal of Geophysical Research: Oceans* 118, no. 4 (2013): 1759-1774.

[RBGV08] Rautek, Peter, Stefan Bruckner, Eduard Gröller, and Ivan Viola. "Illustrative visualization: new technology or useless tautology?." *ACM SIGGRAPH Computer Graphics* 42, no. 3 (2008): 4.

[RC94] Rao, Ramana, and Stuart K. Card. "The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 318-322. ACM, 1994.

[RE01] Rheingans, Penny, and David Ebert. "Volume illustration: Nonphotorealistic rendering of volume models." *IEEE Transactions on Visualization and Computer Graphics* 7, no. 3 (2001): 253-264.

[RLH14] Radeva, Nadezhda, Lucien Levy, and James Hahn. "Generalized temporal focus+ context framework for improved medical data exploration." Journal of digital imaging 27, no. 2 (2014): 207-219.

[RPS01] Reinders, Freek, Frits H. Post, and Hans JW Spoelder. "Visualization of time-dependent data with feature tracking and event detection." *The Visual Computer* 17, no. 1 (2001): 55-71.

[Rob07] Roberts, Jonathan C. "State of the art: Coordinated & multiple views in exploratory visualization." In Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on, pp. 61-71. IEEE, 2007.

[Rup95] Ruppert, Jim. "A Delaunay refinement algorithm for quality 2-dimensional mesh generation." *Journal of algorithms* 18, no. 3 (1995): 548-585.

[SA82] Spence, Robert, and Mark Apperley. "Data base navigation: an office environment for the professional." Behaviour & Information Technology 1, no. 1 (1982): 43-54.

[SB92] Sarkar, Manojit, and Marc H. Brown. "Graphical fisheye views of graphs." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 83-91. ACM, 1992.

[ECS00] Ellsworth, David, Ling-Jen Chiang, and Han-Wei Shen. "Accelerating time-varying hardware volume rendering using tsp trees and color-based error metrics." In Proceedings of the 2000 IEEE symposium on Volume visualization, pp. 119-128. ACM, 2000.

[SCM99] Shen, Han-Wei, Ling-Jen Chiang, and Kwan-Liu Ma. "A fast volume rendering algorithm for time-varying fields using a time-space partitioning (TSP) tree." In *Proceedings of the conference on Visualization'99: celebrating ten years*, pp. 371-377. IEEE Computer Society Press, 1999.

[SGG14] Streit, Marc, Samuel Gratzl, Michael Gillhofer, Andreas Mayr, Andreas Mitterecker, and Sepp Hochreiter. "Furby: fuzzy force-directed bicluster visualization." BMC bioinformatics 15, no. 6 (2014): S4.

[SGL08] Stasko, John, Carsten Görg, and Zhicheng Liu. "Jigsaw: supporting investigative analysis through interactive visualization." *Information visualization* 7, no. 2 (2008): 118-132.

[Shn96] Shneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualizations." In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pp. 336-343. IEEE, 1996.

[Sim81] Simon, Herbert A. "The sciences of the artificial, (1981)." MIT Press.

[SM03] Shchepetkin, Alexander F., and James C. McWilliams. "A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate." *Journal of Geophysical Research: Oceans* 108, no. C3 (2003).

[SPA15] Samsel, Francesca, Mark Petersen, Gregory Abram, Terece L. Turton, David Rogers, and James Ahrens. "Visualization of ocean currents and eddies in a high-resolution global ocean-climate model." In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis 2015*. 2015.

[Spe14] Spence, Robert. *Information visualization*. 3rd Edition. Springer, 2014.

[SSZC94] Samtaney, Ravi, Deborah Silver, Norman Zabusky, and Jim Cao. "Visualizing features and tracking their evolution." *Computer* 27, no. 7 (1994): 20-27.

[Str10] Stray, Jonathan, "Overview: a tool for exploring large document sets." http://jonathanstray.com/overview-a-tool-for-exploring-large-document-sets, December 1, 2010.

[SW97] Silver, Deborah, and Xin Wang. "Tracking and visualizing turbulent 3d features." *IEEE Transactions on Visualization and Computer Graphics* 3, no. 2 (1997): 129-141.

[SWRG02] Suh, Bongwon, Allison Woodruff, Ruth Rosenholtz, and Alyssa Glass. "Popout prism: adding perceptual principles to overview+ detail document interfaces." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 251-258. ACM, 2002.

[Symplicity] Symplicity, http://symplicity.com.

[TAHS06] Tominski, Christian, James Abello, Frank Van Ham, and Heidrun Schumann. "Fisheye tree views and lenses for graph visualization." In Information Visualization, 2006. IV 2006. Tenth International Conference on, pp. 17-24. IEEE, 2006.

[TGK14] Tominski, Christian, Stefan Gladisch, Ulrike Kister, Raimund Dachselt, and Heidrun Schumann. "A survey on interactive lenses in visualization." EuroVis State-of-the-Art Reports 3 (2014).

[TGK17] Tominski, Christian, Stefan Gladisch, Ulrike Kister, Raimund Dachselt, and Heidrun Schumann. "Interactive lenses for visualization: An extended survey." In Computer Graphics Forum, vol. 36, no. 6, pp. 173-200. 2017.

[THP12] Takle, Jay, Katrin Heitmann, Tom Peterka, Deborah Silver, George Zagaris, and Salman Habib. "Tracking and Visualizing Evolution of the Universe: In Situ Parallel Dark Matter Halo Merger Trees." In High Performance Computing, Networking, Storage and Analysis (SCC), 2012 SC Companion:, pp. 1482-1483. IEEE, 2012.

[TS02] Tory, Melanie, and Colin Swindells. Exovis: An overview and detail technique for volumes. Technical Report SFU-CMPTTR2002-05, Computing Science Dept., Simon Fraser University, 2002.

[TSAA12] Tominski, Christian, Heidrun Schumann, Gennady Andrienko, and Natalia Andrienko. "Stacking-based visualization of trajectory attribute data." IEEE Transactions on visualization and Computer Graphics 18, no. 12 (2012): 2565-2574.

[UCB14] US Census Bureau, Where do college graduates work? https://www.census.gov/dataviz/visualizations/stem/stemhtml/, July 2014.

[VFSG06] Viola, Ivan, Miquel Feixas, Mateu Sbert, and Meister Eduard Groller. "Importance-driven focus of attention." IEEE Transactions on Visualization and Computer Graphics 12, no. 5 (2006).

[VG05] Viola, Ivan, and Eduard Gröller. "Smart visibility in visualization." In *Computational Aesthetics*, pp. 209-216. 2005.

[VKG05] Viola, Ivan, Armin Kanitsar, and M. Eduard Groller. "Importance-driven feature enhancement in volume visualization." *IEEE Transactions on Visualization and Computer Graphics* 11, no. 4 (2005): 408-418.

[VW14] Van den Elzen, Stef, and Jarke J. Van Wijk. "Multivariate network exploration and presentation: From detail to overview via selections and aggregations." IEEE Transactions on Visualization and Computer Graphics 20, no. 12 (2014): 2310-2319.

[VZS18] Valdez, Andre Calero, Martina Ziefle, and Michael Sedlmair. "Priming and anchoring effects in visualization." IEEE Transactions on Visualization & Computer Graphics 1 (2018): 584-594.

[Wei91] Weiss, John. "The dynamics of enstrophy transfer in two-dimensional hydrodynamics." *Physica D: Nonlinear Phenomena* 48, no. 2-3 (1991): 273-294.

[WH07] Wohlfart, Michael, and Helwig Hauser. "Story telling aspects in volume visualization." In *Proceedings of EuroVis*, pp. 91-98. 2007.

[WHP11] Williams, Sean, Matthew Hecht, Mark Petersen, Richard Strelitz, Mathew Maltrud, J. Ahrens, Mario Hlawitschka, and Bernd Hamann. "Visualization and analysis of eddies in a global ocean simulation." In *Computer Graphics Forum*, vol. 30, no. 3, pp. 991-1000. Blackwell Publishing Ltd, 2011.

[WL90] Wehrend, Stephen, and Clayton Lewis. "A problem-oriented classification of visualization techniques." In Visualization, 1990. Visualization'90., Proceedings of the First IEEE Conference on, pp. 139-143. IEEE, 1990.

[WPS16] Woodring, Jonathan, Mark Petersen, Andre Schmeiβer, John Patchett, James Ahrens, and Hans Hagen. "In situ eddy analysis in a high-resolution ocean climate model." *IEEE transactions on visualization and computer graphics* 22, no. 1 (2016): 857-866.

[WWDW06] Wang, Weixin, Hui Wang, Guozhong Dai, and Hongan Wang. "Visualization of large hierarchical data by circle packing." In Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 517-520. ACM, 2006.

[WYM08] Wang, Chaoli, Hongfeng Yu, and Kwan-Liu Ma. "Importance-driven time-varying data visualization." IEEE Transactions on Visualization and Computer Graphics 14, no. 6 (2008): 1547-1554.

[WZMK05] Wang, Lujin, Ye Zhao, Klaus Mueller, and Arie Kaufman. "The magic volume lens: An interactive focus+ context technique for volume rendering." In Visualization, 2005. VIS 05. IEEE, pp. 367-374. IEEE, 2005.

[ZCPB11] Zhao, Jian, Fanny Chevalier, Emmanuel Pietriga, and Ravin Balakrishnan. "Exploratory analysis of time-series with chronolenses." IEEE Transactions on Visualization and Computer Graphics 17, no. 12 (2011): 2422-2431.

[ZCEV17] Zhang, C., M. W. A. Caan, T. Höllt, E. Eisemann, and A. Vilanova. "Overview+ Detail Visualization for Ensembles of Diffusion Tensors." In Computer Graphics Forum, vol. 36, no. 3, pp. 121-132. 2017.

[ZM95] Zhu, Zhifan, and Robert J. Moorhead. "Extracting and visualizing ocean eddies in time-varying flow fields." In *Proceedings of the 7th International Conference on Flow Visualization*, pp. 206-211. Seattle, WA, 1995.

Appendix A: Implementation of Case Study 1

A.1: Computation Performance

The implementation of the eddy computation program was performed on a Dell XPS desktop with an Intel Core i7-4770 (3.40 GHz * 8) CPU, 16 GB memory, and a NVidia GTX 760 graphics card (2 GB video memory), running under Ubuntu 16.04 LTS.

The 1095 ROMS data frames (1.78 TB in total) were downloaded from the data server and were stored in a local folder. Some important parameters of the program running include a) depths of eddy features≥100 meters; b) volume of eddy features≥50 data points; c) duration of eddy paths≥30 days. The performance of the computation is evaluated by elapsed times and the sizes of the output files. In the program, three timer functions were set to count the elapsed times of eddy extraction (include data reading and preprocessing), eddy tracking, and eddy path isolation. The generated files were saved in a different folder from the raw data files. Table A.1 lists the elapse times and the file sizes of the output (the tracking table and eddy path list are negligible in file size and so are not listed in the table)

Table A.1. Elapsed Times and Sizes of Output Files

Elapsed Times			**File Sizes of the Output Files**		
Extraction	Tracking	Isolation	*.day files	*.uocd files	*.path files
2 h 46 min	2 h 33 min	5 h 41 min	10.52 GB	8.67 GB	225.67 MB

A.2: Eddy Computation Program

More information about the eddy computation program and the accesses to the source code and the data can be found at:

http://vizlab.rutgers.edu/eddyprogram

A.3: Formats of the Output Files

Only *.day files and *.path files are used for visualization.

***.uocd file**

```
******uocd******            // file header
number of eddy features
eddy ID
volume  c_x  c_y  c_z       // volume and centroids
point ID  x0  y0  z0
point ID  x1  y1  z1
point ID  x2  y2  z2
point ID  x3  y3  z3
   .
   .
   .
```

***.day file**

******day******

date

155 155 155 // signifies the start of this eddy feature

spin direction

volume c_x c_y c_z l_x l_y l_z u_x u_y u_z

number of nodes min_temp max_temp min_salt max_salt min_speed max_speed

x0 y0 z0 temp0 salt0 speed0

x1 y1 z1 temp1 salt1 speed1

x2 y2 z2 temp2 salt2 speed2

x3 y3 z3 temp3 salt3 speed3

x4 y4 z4 temp4 salt4 speed4

.

.

.

number of connections

3 vertex ID vertex ID vertex ID

3 vertex ID vertex ID vertex ID

3 vertex ID vertex ID vertex ID

.

.

.

0 // signifies the end of this eddy feature

155 155 155

spin direction

.

.

.

***.trackTable file**

Time Step #

eddy ID at previous time step -1 corresponding eddy ID at current time step

.

.

.

Here the “-1” acts as a delimiter. If an eddy from previous time step disappear at current time step, there is nothing to the right of “-1”. If a new eddy appears at current time step, then the left side of “-1” will be empty.

***.sequence file**

Eddy ID Time Step # Event Eddy ID Time Step # Event Eddy ID Time Step # Event **……**

.

.

.

Here, the numbers 1 to 5 are assigned to represent the events birth, continuation, merge, split, and death in the eddy evolution process.

***.path file**

******path******

min_temp max_temp min_salt max_salt min_speed max_speed //physical properties of this eddy path

date period

155 155 155 // signifies the start of this time step

spin direction

volume c_x c_y c_z l_x l_y l_z u_x u_y u_z

number of nodes min_temp max_temp min_salt max_salt min_speed max_speed // properties of this time step

x0 y0 z0 temp0 salt0 speed0

x1 y1 z1 temp1 salt1 speed1

x2 y2 z2 temp2 salt2 speed2

x3 y3 z3 temp3 salt3 speed3

x4 y4 z4 temp4 salt4 speed4

.

.

.

number of connections

3 vertex ID vertex ID vertex ID

3 vertex ID vertex ID vertex ID

3 vertex ID vertex ID vertex ID

.

.

.

0 // signifies the end of this time step

A.4: Accesses to the WebGL Eddy Visual Explorer and the Data for Visualization

A link to the WebGL Eddy Visual Explorer:

http://vizlab.rutgers.edu/eddyviz.html

Please note, the number of opened tabs or webpages in a browser will affect the speed of data loading and visualization rendering.

Eddy *.path files can be downloaded at ***here***

Eddy *.day files can be downloaded at ***here***

Appendix B: Implementation of Case Study 2

B.1: Scalability Analysis

The scalability of the JobViz tool in the second case study is measured by how many job postings it can handle. As described before, listing the actual job posting data around the arc quickly became unfeasible and was replaced by a hierarchical/group view of the job postings. While job categories are the groupings used in this work, any other groupings can also be used. As the number of job postings increase, although the screen size doesn't set any limit to display the data, the responding time to compute on-the-fly the bundle lines and tag cloud increases. We tested the scalability of JobViz system based on the time it needs to finish three typical tasks: loading the system, searching by major, and searching by job. For both searching by major and by job, we define the cost time as the elapsed time from clicking a node to displaying the major-job bundle-lines and the tag cloud in the screen.

Our experiments were conducted on a Mac Pro (2.5 GHz Intel Core i7, 16 GB Memory, and Mac OS X EI Capitan 10.11.3) installed the latest Firefox web browser. In the experiment, we record the elapsing time of each task for data files with different number of job postings from 1000 – 10, 000. The cost time is calculated by the difference between the date instances before and after each task. Each task is repeated five times and we take the average of its cost time. For the task "search by major", we chose "All Major" to test because it links to the largest number of jobs and job categories. And for the task "search by job", we chose the job category "Marketing/Sales" because it contains more jobs than any other job category does. The time costs of the three tasks are represented as line chart (blue line for “load system”; orange line for “search by major”; gray line for “search by job”) in Figure B.1.

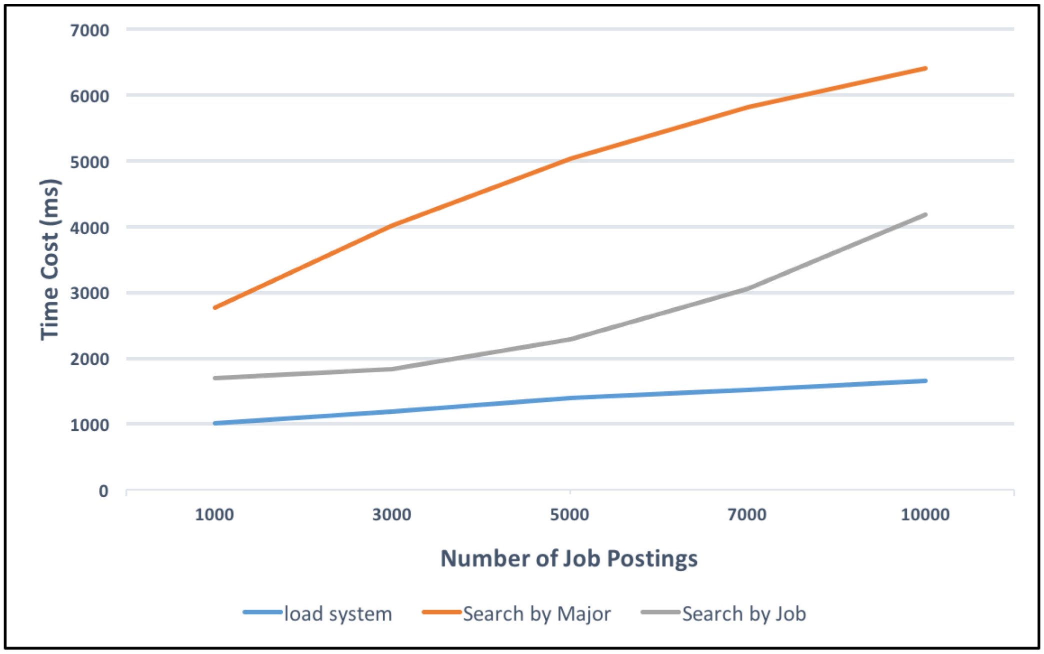


Figure B.1: Scalability evaluation of JobViz. The X-axis is the number of job postings and the Y-axis is the time cost in millions of seconds.

The results show that search by major costs the most time (up to more than 6 seconds when the number of job postings reaches 10, 000) and that all tasks can be finished within 7 seconds. This suggests that the JobViz can handle more than 10, 000 job postings if we set the maximum delay as 10 seconds.

B.2: Source Code

https://github.com/liuli4016/jobviz_CGA

B.3 License

For information about using or licensing JobViz please refer to the ***Rutgers Copyright Software Licensing Site.***

Appendix C: Accompanying Videos

The accompanying videos to the two case studies can be accessed through the links below.

Case Study 1:

https://drive.google.com/file/d/1uMK11SRV7Qtn5-_gpcTMIHF7K9dp7XmQ/view?usp=sharing

Case Study 2:

https://drive.google.com/file/d/0BwfHRoSTgAdnc3p0WTg5REh3bnM/view?usp=sharing