Designing a PCA-based Collaborative Visual Analytics System

Dong Hyun Jeong, William Ribarsky, and Remco Chang

Abstract—In visual analytics, collaboration is viewed as a knowledge sharing process that helps people perform analytical reasoning tasks effectively. In this paper, we present a collaborative visual analytics tool, iPCA-CE, that supports interactive data analysis using principal component analysis (PCA) on a tabletop display. We define three data analysis scenarios that are addressed when designing the collaborative data analysis system. With the system, users are able to collaboratively analyze data, share ideas or knowledge, and divide their work-load.

Index Terms—Collaborative data analysis, Touch-table, Multi-touch interaction.

1 Introduction

In knowledge management literature, socialization is defined as a process in which people communicate with each other in order to share their ideas or personalized (tacit) knowledge [11, 18]. In visualization, this is achieved through the collaboration process of sharing knowledge, learning, and building consensus through the use of computers [22]. Several researchers have studied users' behavior through collaborative environments in order to better understand this knowledge-sharing process. Mark and Kobsa [17] performed an empirical study to understand the differences between group and individual behavior within collaborative information visualization environments. They found that a group solves the given questions more accurately and spends less time doing so. However, it is still unknown what features should be supported within a collaborative data analysis system on a touch-table in order to reliably gain these benefits.

Analyzing data is a complicated task. If people can combine their efforts in an analytical task, they might have a better chance of solving complex problems or finding obscured information. In this paper, we focus on designing a collaborative visual analytics environment to support interactive data analysis on a touch-table. Since previous research shows that with a more user-friendly collaborative visualization system, people find results more easily and accurately [17], we choose our existing visual analytics system (called iPCA - interactive principal component analysis) and extend it to work on a multi-touch tabletop display. We named the extended version of iPCA as iPCA-CE (interactive PCA within collaborative environments). When designing the collaborative visual analytics system, we carefully consider addressing three different types of collaborative data analysis scenarios (see Section 3 for detail).

The rest of this paper consists of four sections. First we discuss related research in collaborative visualization environments. Then we explain our system's interface design and multi-touch interactions. In section 3, we introduce three collaborative data analysis scenarios supported by our system, and conclude with discussion and future work.

2 Previous Work

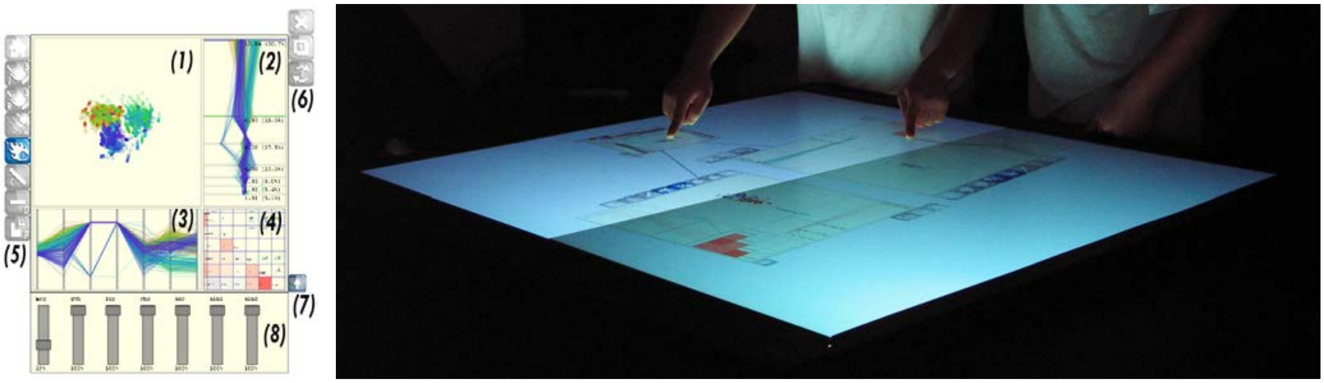
In the past, many notable studies have been done in collaborative visualization. There are roughly three main research trends: building collaborative visualization environments, sharing knowledge through web-based collaborative workspaces, and interactively sharing tacit knowledge with people on a touch surface. In this section, we introduce some of the existing literature.

- *Dong Hyun Jeong is with Charlotte Visualization Center at UNC Charlotte, E-mail: dhjeong@uncc.edu.*
- *William Ribarsky is with Charlotte Visualization Center at UNC Charlotte, E-mail: ribarsky@uncc.edu.*
- *Remco Chang is with Charlotte Visualization Center at UNC Charlotte, E-mail: rchang@uncc.edu*

Collaboration has been described as the process of sharing tacit knowledge between people [18]. Mark and Kobsa [17] defined collaborative information visualization behavior as a social process as well as a cognitive process because it involves both interpreting visualization and coordinating complex social activities. Although the knowledge sharing process and the cognitive process are both broadly regarded as important research topics [22], limited study has been done in visualization. However, building collaborative visualization environments has a long history [5, 13]. Coleman et al. [5] provided four general reasons why collaborative visualization is compelling. (1) Experts' knowledge can be available any time and at any place. (2) The expertise can be transferred to others, improving the local level of knowledge. (3) Based on the supported accessibility, visualization products can be reviewed and modified as they are produced, reducing turn-around time. (4) Remote accessibility also reduces the need to relocate the expertise physically. Johnson [13] defined collaborative visualization as a subset of computer-supported cooperative work (CSCW) in which control over parameters or products of the scientific visualization process is shared.

More recently, Grimstead et al. [8] reviewed 42 collaborative visualization systems in terms of five attributes: number of simultaneous users, user access control, communication architecture, type of transmitted data, and user synchronization. They found that the synchronous system has the benefits of bringing groups of individuals together over a distance, bridging the knowledge gaps among them, and building their knowledge structure concurrently. But, they noticed that the synchronous system is still limited in that people have to be in front of computer machines at the same time. However, in an asynchronous collaborative visualization system, collaboration occurs at different times. If people are in different time zones and different places, an asynchronous collaborative system might be beneficial [16]. Once important knowledge is found, it can be shared with others asynchronously at their own convenience. However, it is still unclear how collaborative visualization should be designed. Because of this, Heer and Agrawala [9] provide design considerations for asynchronous collaboration in visual analytics environments. Ma [15] noted that sharing visualization resources will provide the eventual support for a collaborative workspace. He discussed existing web-based collaborative workspaces in terms of sharing high-performance visualization facilities, visualizations, and findings. Burkhard proposed a collaboration process of transferring knowledge between at least two persons or group of persons [2].

Although much research has been done in collaborative visualization, there has been less work in collaboration on touch surfaces. Isenberg and Fisher [10] designed a system (called Cambiera) to support collaborative search through large text document collections on a touch surface. They considered collaborative activities to involve not just searching through documents, but also building individual's findings and maintaining awareness of another person's work. North et al. [19] studied how users approach a multi-touch interface and what types of gestures they are willing to use. In the study, they performed object manipulation tasks on a physical table, a multi-touch table, and



(a) (b)

Fig. 1. The system overview (a) showing four views ($1 \sim 4$), several selectable buttons ($5 \sim 7$), and a set of sliderbars (8) with E.Coli dataset (336×7 matrix). The system supports changing the scale and the location as well as manipulating the projected data item(s). With using the system, multiple people can collaborate each other on a multi-touch tabletop display interactively (b).

a desktop computer. From the study, they found that people completed the tasks significantly faster with multi-touch interactions on a multi-touch table than with mouse interactions on a desktop computer. Furthermore, they found that subjects are significantly faster on a physical surface than on a touch surface.

However, to the best of our knowledge on collaborative visualization, what we should support when designing a collaborative data analysis system has not been broadly studied. Developing this design philosophy requires understanding how people act on a touch table, especially when analyzing data. In the following sections, we will provide a detailed explanation about how our system is designed and what features are supported.

3 Collaborative Visual Analytics Application

Understanding how people act when analyzing data is an important research topic in visual analytics, but it is also extremely challenging [22]. In this paper, we focus our research to the understanding of analytical behavior to be strictly within the context of a collaborative environment. Our work begins with designing a useful collaborative visual analytics application with which people can easily share their ideas and knowledge. Our design philosophy is not to develop a new visual analytics application, but instead we extend an already known and useful visual analytics application to work on a touch table. Specifically, we choose our existing visual analytics application (iPCA) because studies have shown that user-friendly visualizations in a collaborative environment enable users to find results more accurately [17]. In our previous study [12], we found that iPCA is both easy to use and effective in helping users learn about PCA and the datasets they are using.

Data analysis is often considered as a stand-alone analytical task. However, as previous research has shown, analysis of (empirical) data in collaborative environments is important and should be considered while developing visualization applications [4, 7]. While collaborative analytics can occur in different interaction modalities, we focus specifically on collaboration on a multi-touch table based on existing work that demonstrated potential increase in analysis performance [10].

3.1 System Design

iPCA is designed to help the user understand the complex black box operation of Principal Component Analysis [14] and interactively analyze data [12]. We extend this application to support collaborative data analysis on a touch table.

Figure 1(a) shows the system overview, which includes four views, touchable buttons, and a set of dimension sliders. The overall interface is developed with OpenGL. It supports multiple-touch interactions on a horizontal display. The multi-touch display system was designed at the Renaissance Computing Institute (RENCI) [1]. It provides a 62" diagonal work surface (42 x 46), in which two HD resolution projection displays create images on the surface to support multiple people working together. Figure 1(b) represents the overall workspace, in which two people are collaborating on a touch table.

Like the original iPCA, our extended application (iPCA-CE) consists of four views: Projection view (Figure 1(a-1), Eigenvector view (Figure 1(a-2), Data view (Figure 1(a-3), and Correlation view (Figure 1(a-4)). In the Projection view, all data items are projected based on the first and second principal components by default. The Eigenvector view displays the calculated eigenvectors and eigenvalues in a vertically projected parallel coordinate. The distances between the eigenvectors in the parallel coordinate view vary based on their eigenvalues, separating the eigenvectors based on their mathematical weights. The Data view shows the original data points in a parallel coordinate. The Correlation view represents Pearson-correlation coefficients and relationships between variables as a matrix of scatter plots and values. All views are closely connected, so that an action in one view can affect the other views. If the user interactively changes the elements in one view, its corresponding results are updated in other views (*brushing & linking*). This interactivity thus allows the user to infer relationships between the coordinated spaces (see [12] for detail).

There are a total of 12 touchable buttons designed: 8 buttons are for interacting with represented data items (Figure 1(a-5)), 3 buttons are for controlling the application (Figure 1(a-6)), and the last button (Figure 1(a-7)) is for making the sliderbars appear and disappear. Table 1 represents the touchable buttons and their meanings.

The system supports basic multi-touch operations such as zooming,

Table 1. Touchable buttons and their meanings

Button	Meaning	Button	Meaning
	Go back to the initial state		Delete the selected item(s)
	Individual item selection		Partition the selected item(s) into a new workspace
	Range item(s) selection		Close the application
	Manipulation		Create a new application
	Trail enable – on/ off		Rotate the application
	Cancel the selected item(s)		Make the sliderbar panel appear / disappear

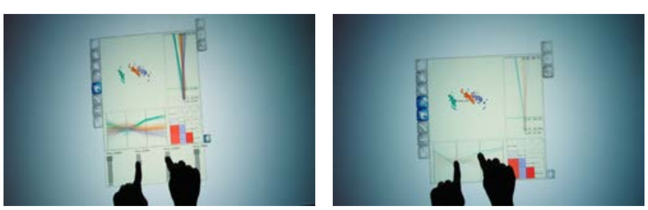
panning, and rotation. The zooming operation is activated by making two finger touches closer (zoom-in) and farther apart (zoom-out). The panning operation is initiated by dragging a finger on the surface. However, the rotation only works when the rotation option (a touchable button) is enabled. We adopt this passive operation because if the user accidently changes the angle between two touches during analysis, the rotation operation is activated unintentionally, and sometimes distracts people from concentrating on analyzing the data.

Additionally, the system provides several data operations such as individual item selection, range item(s) selection, deletion, and manipulation. Both the individual item selection and the range item(s) selection operations are allowed in all four views. In Data View and Eigenvector View, where the visualizations are parallel coordinates, selection means clicking on a single line or brushing a range of items. In Projection View and Correlation View, the user can either click on a single dot or draw an enclosed space upon which all data items within the space will be selected. In analysis using PCA, a common task is for the user to remove outliers. The deletion operation is to remove the selected data item(s) from the PCA calculation. The manipulation is the operation, which allows the user to see the relationship between principal component(s) and data dimensions.

3.2 Multi-touch Interactions

As shown in Figure 1(a), the E.coli dataset has 7 dimensional attributes. But it is not linearly separable by a PCA calculation since PCA assumes that the input data are always linear. Because of this, weighted principal component analysis (WPCA) is often considered, which allows different weights on different variables as $s_1, s_2, ..., s_n$ [14]. This approach assumes that data are not always linearly increasing or decreasing, and there may be reason to allow different observations to have different weights. To provide the ability to analyze the data non-linearly, iPCA has a set of dimension sliderbars, which allow the user to change the dimension contributions of each dimension. However, with a mouse-based interface, the user has to try all possible combinations of dimension contribution changes with a series of single mouse inputs to fully understand and analyze the data. iPCA-CE gives the user the ability to change several dimensions at once on a multi-touch table, thus permitting much more effective exploration of the high dimensional space and how the dimensions correlate. Figure 2(a) shows an example in which the user changes dimension contributions by moving the sliderbars with two finger touches.

iPCA-CE also allows the user to alter the values of data items. For instance, if the user drags a data item in the Projection View towards the positive direction along the x-axis (increasing the data point's value in the first principle component), the user should be able to immediately observe in the Data View how that change affects the values of that data item in the original data space, thus shedding light on the relationship between the first principle component and all dimensions in the original data space. Figure 2(b) shows the user manipulating the selected data item in the Data view with two finger touches.



(a) (b)

Fig. 2. Multi-touch interactions. (a) The user changes the dimension contributions using sliderbars and (b) the user directly modifies the values of a data item in the Data view.

4 Collaborative Data Analysis

A collaboration process can occur through the use of collaborative visual environments. However, the most natural method for sharing tacit knowledge is still direct communication between users. In either case, the users are actively sharing their discoveries and tacit knowledge and incorporating each other's domain expertise into their own. However, understanding and addressing analytical procedures are important when designing a useful collaborative visual analytics application. In general, collaborative environments on a touch table support either tightly coupled collaboration (having a shared workspace and working together) or loosely coupled collaboration (having independent workspaces and working alone for long periods of time) [21]. In our collaborative visual analytics application, we considered addressing three types of analytical scenarios: 1) people are collaborating with others by looking at the same results (tightly coupled collaboration); 2) people are analyzing the same dataset with their own individual workspaces (loosely coupled collaboration); and 3) people are working with a partitioned dataset within their own workspaces (tightly and loosely coupled collaboration). The third scenario, however, is especially important because it supports both tightly and loosely coupled collaboration (see Section 3.3 for detail).

4.1 Looking at the Same Results

In visual analytics, people are often working together by looking at the same results displayed on a screen, which is a common analytical procedure when collaborating with others. Most visual analytics applications support this analytical procedure, as it works in any types of display system. However, on a touch table, existing visual analytics applications allow multiple people to work at the same time (tightly coupled collaboration). Butkiewicz et al. [3] designed a geospatial analysis tool running on a touch table, with which people can interactively create multiple probes based on their regions of interest. In such an environment, people can easily share ideas, findings, and their expertise with others by looking at the same results. This is also somewhat related to a learning system, in which an expert explains interesting results or his personalized knowledge to novice users so they can come up with solutions and analyze the data effectively on their own.

Figure 3(a) shows two users working together by looking at and interacting with the same representation displayed on a touch surface. In this example, the user (left) is trying to show the effectiveness of data value changes to the other user (right). Within this environment, users can directly communicate with each other focusing on the same visual representation and results.

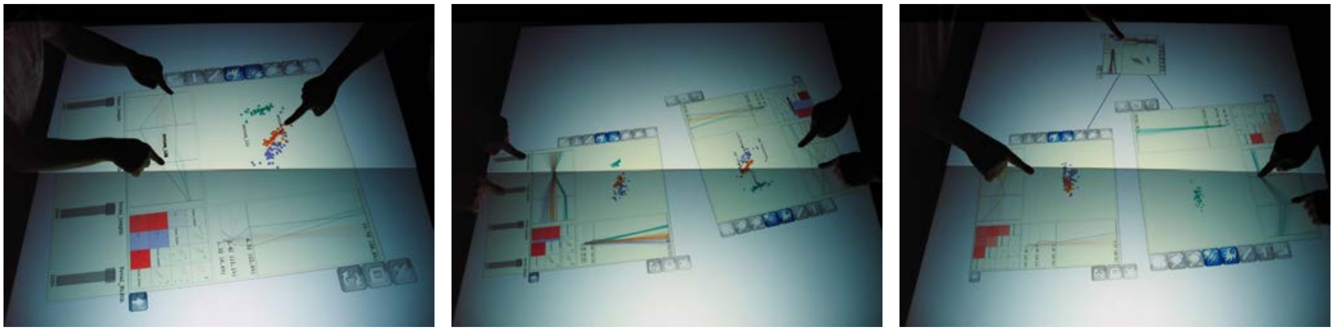
4.2 Working with the Same Dataset

In collaborative visualization applications, a common analytical procedure is to work with the same dataset synchronously and asynchronously. Because of this, most existing collaborative visualization applications support both synchronous and asynchronous knowledge sharing. However, in our collaborative visual analytics application, we only consider synchronous collaboration.

On a multi-touch table, people can analyze the dataset by looking at different representations. Once a person finds an interesting result, he can directly communicate it by passing or showing the result to a colleague. This is somewhat related to the analytical procedure described in Section 3.1. However, having individual workspaces may increase the overall performance of finding hidden information and analyzing the data (loosely coupled collaboration). Figure 3(b) shows an example in which people collaboratively analyze the public Iris dataset with their own workspaces. In Figure 3(b), the user (left) analyzes the data by changing the dimension contribution of the first (Sepal length) and second (Sepal width) variables, and the other user (right) manipulates the values of the selected data item in the parallel coordinates within the Data view to understand how the selected data item(s) are placed in a certain cluster.

4.3 Working with the Partitioned Datasets

In data analysis, data partitioning is an important pre-processing operation. For instance, a Bayesian phylogenetic analysis tool (MrBayes 3 [20]) partitions data according to the data type by default, and then analyzes the partitioned datasets separately. This is because most real-world datasets do not exist in the form of a combined dataset. Also,



(a) (b) (c)

Fig. 3. The pictures show people performing multiple collaborative data analysis scenarios in iPCA-CE system with the Iris dataset (140×4 matrix). (a) People are working together by looking at the same tool and results, (b) working with the same dataset, but in different workspaces, and (c) working with partitioned datasets in their own workspaces. The lines between workspaces in (c) indicate the independent workspaces (the partitioned datasets (left and right)) from the shared workspace (the original dataset (top middle)).

people often tend to focus on analyzing a specific dataset based on their interests or personalized (tacit) knowledge. In financial fraud analysis, analysts tend to investigate specific financial datasets (e.g. the transactions between two specific countries) based on their experience [6].

In iPCA-CE, users are able to interactively partition the dataset in order to collaborate with others. Once the dataset is partitioned, the partitioned dataset creates a (blue) connected line to its original dataset. In this analytical scenario, the system supports both loosely coupled collaboration and tightly coupled collaboration. The system is designed to support creating multiple independent workspaces from a shared workspace. Figure 3(c) shows a shared work space and two independent workspaces. The unpartitioned dataset is projected in the shared workspace and the partitioned datasets are displayed in the independent workspaces.

5 Conclusion and Future work

Since data analysis is a complex analytical task, many useful visual analytics applications are designed to assist users analyzing data effectively. However, limited research has been done on understanding how to support data analysis on a touch table. In this paper, we described three important analytical scenarios that should be supported when designing a collaborative data analysis application on a touch table. We also designed a collaborative data analysis application (iPCA-CE) based on these analytical scenarios.

Since how people share ideas or personalized (tacit) knowledge on a touch-table when solving complex analytical tasks is still not known, our future work includes understanding the human knowledge sharing process on a touch table.

Acknowledgements

The authors wish to thank Caroline Ziemkiewicz for her help on revising the drafts.

References

[1] Renaissance computing institute, 2009. http://www.renci.org.

[2] R. A. Burkhard. Learning from architects: The difference between knowledge visualization and information visualization. In *IV '04: Proceedings of the Information Visualisation, Eighth International Conference*, pages 519–524, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[3] T. Butkiewicz, D. H. Jeong, W. Ribarsky, and R. Chang. Hierarchical multi-touch selection techniques for collaborative geospatial analysis. In *SPIE Defense, Security, and Sensing 2009*, 2009.

[4] I. Carlbom, W. M. Hsu, G. Klinker, R. Szeliski, K. Waters, M. Doyle, J. Gettys, K. M. Harris, T. M. Levergood, R. Palmer, L. Palmer, M. Picart, D. Terzopoulos, D. Tonnesen, M. Vannier, and G. Wallace. Modeling and analysis of empirical data in collaborative environments. *Commun. ACM*, 35(6):74–84, 1992.

[5] J. Coleman, A. Goettsch, A. Savchenko, H. Kollmann, W. Kui, E. Klement, and P. Bono. Teleinvivotm: towards collaborative volume visualization environments. *IEEE Comput. Graph. Appl.*, 20(6):801–811, 1996.

[6] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *IEEE Comput. Graph. Appl.*, 29(3):52–61, 2009.

[7] J. Ekanayake, S. Pallickara, and G. Fox. A collaborative framework for scientific data analysis and visualization. In *Collaborative Technologies and Systems, 2008. CTS 2008. International Symposium on*, pages 339–346, May 2008.

[8] I. J. Grimstead, D. W. Walker, and N. J. Avis. Collaborative visualization: A review and taxonomy. *Distributed Simulation and Real-Time Applications, IEEE International Symposium on*, 0:61–69, 2005.

[9] J. Heer and M. Agrawala. Design considerations for collaborative visual analytics. *Information Visualization*, 7(1):49–62, 2008.

[10] P. Isenberg and D. Fisher. Collaborative brushing and linking for colocated visual analytics of document collections. *Computer Graphics Forum (EuroVis 2009)*, 28(3):1031–1038, 2009.

[11] D. H. Jeong, R. Chang, and W. Ribarsky. An alternative definition and model for knowledge visualization. IEEE Visualization 2008 Workshop on Knowledge Assisted Visualization, 2008.

[12] D. H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang. ipca: An interactive system for pca-based visual analytics. *Computer Graphics Forum (EuroVis 2009)*, 28(3):767–774, 2009.

[13] G. Johnson. Collaborative visualization 101. *ACM SIGGRAPH - Computer Graphics*, 32(2):8–11, 1998.

[14] I. Jolliffe. *Principal Component Analysis*. Springer, second edition, 2002.

[15] K.-L. Ma. Creating a collaborative space to share data, visualization, and knowledge. *SIGGRAPH Comput. Graph.*, 41(4):1–4, 2007.

[16] F. T. Marchese and N. Brajkovska. Fostering asynchronous collaborative visualization. In *IV '07: Proceedings of the 11th International Conference Information Visualization*, pages 185–190, Washington, DC, USA, 2007. IEEE Computer Society.

[17] G. Mark and A. Kobsa. The effects of collaboration and system transparency on cive usage: An empirical study and model. *Presence*, 14:2005, 2005.

[18] I. Nonaka and H. Takeuchi. *The Knowledge Creating Company*. Oxford University Press, 1995.

[19] C. North, T. Dwyer, B. Lee, D. Fisher, P. Isenberg, G. Robertson, and K. Quinn. Understanding multi-touch manipulation for surface computing. In *In: Proceedings of 12th IFIP TC13 Conference in Human-Computer Interaction (Interact 2009)*, pages 236–249. Springer, 2009.

[20] F. Ronquist and J. P. Huelsenbeck. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572–1574, 2003.

[21] A. Tang, M. Tory, B. A. Po, P. Neumann, and M. S. T. Carpendale. Collaborative coupling over tabletop displays. In *CHI*, pages 1181–1190, 2006.

[22] J. J. Thomas and K. A. Cook. A visual analytics agenda. *IEEE Comput. Graph. Appl.*, 26(1):10–13, 2006.