Statement of Purpose

who I am and what I want to do

Fumeng Yang

March 6, 2016

Contents

1 Research Statement

1.1 Intended Research

Data visualization is becoming a crucial approach to rapidly interpret enormous amounts of information. From Confirmatory Data Analysis to Exploratory Data Analysis, data visualization gives people the power to interact with a dataset and find the story behind it. However, with few exceptions, data visualization is considered to be an applied discipline under few theoretical foundations (i.e. color theory [1], data-visual mapping [5], and "best practice" design principles [6]). Without establishing the necessary theoretical foundations, a data visualization technology usually fails in being scaled to a different application domain. Furthermore, in the days of big data, it continues to fail in addressing limitations in both human cognition and computational resources.

Facing these issues, my research goal is to develop theoretical foundations for visualization and apply them in practice. While this is a long-term goal that has various topics, in particular, my Ph.D. topic would be **Model-Driven Approaches for Visualization: Perception, Design and Computation**, where perceptual models of visualization would be first built then applied in design and computation.

The outcome of this research has both theoretical and practical applications to data visualization and beyond. For visualization researchers and practitioners, integrating perceptual models with computation provides an approach to do approximate computing, which can address the limitations in human cognition and computational resources.

More broadly, these models have an impact on other disciplines related to the science of big data such as databases and machine learning. For example, the perceptual models can be applied to perception-based sampling, progressive visualization and active learning.

The model-driven approaches for visualization could impact how people think of visualization in a variety of fields. Visualization should not be just an outcome that is built at the end of a chain of science or industry, but should also be integrated into the process of design and production.

1.2 Completed Research and Approach

My previous work supports a model-driven approach for visualization.

One project, which published in *IEEE Transactions on Visualization and Computer Graphics* and presented at *IEEE InfoVis* conference in 2014, is the perceptual models of correlation in visualizations [4]. This project hypothesized that it is possible to quantitatively compare and rank visualizations without exhaustive empirical testing that is common in the field. In this project, the perception of correlation in nine commonly-used visualizations is measured by perceptual discrimination. Perceptual discrimination in all of these visualizations was found to be a linear function of correlation, which provides the possibility to build comparable perceptual models. These perceptual models draw a baseline to develop visualization theory based on quantitatively modeling, while the models themselves apply to practice. For example, at the end of the paper, using the ranking of effectiveness of these visualizations, designers are able to choose the right one for correlation accordingly.

However, perceptual models only show results but don't take into account the gap between visual encoding and human perception. This issue not only limits the power and potential applications of the models but also prevents researchers from building comprehensive visualization theory. In the later work [7], which is prepared for *IEEE InfoVis 2016*, the bridge between visual encoding and human perception is hypothesized to be a visual feature which is used as the proxy in judging correlation, namely in scatterplots. In order to find the evidence, forty-four visual features were collected from the literature in both perceptual psychology and visualization, and further evaluated based on the accuracy of predicting performance in the experiment. As a result, two visual features were found to be the target. Thus, the perceptual model in the previous work [4] was explained as visual encoding creates visual features which was perceived by humans. This project builds a piece of theoretical foundations for visualization based on perceptual psychology and other relevant fields. The approach, that is to combine computational modeling with perceptual experiments, is able to address some fundamental problems in the visualization field such as how visualization is dis-coded by humans and how this knowledge applies to the design and evaluation of visualization.

Studying scatterplots is a first step for building a comprehensive perceptual model for correlation that scales widely. At the same time, it is essential to study other perceptual models. Furthermore, the computational methods should also be applied to quantitatively measuring other aspects of visualization (i.e. design). Eventually my goal is to integrate visualization systems with models in practice to address the limitations in human cognition and computational resources.

1.3 Prospective Projects

Under the topic of model-driven approaches for perception of visualization, a couple of prospective projects have been shaped based on my previous work. At the same time, having a large goal of research, other prospective projects are also stated as following.

- Apply well-established perceptual models

  Except building new perceptual models for *comprehensive level* [2] concepts such as correlation, some well-established perceptual models, such as pixel-level perceptual models [3, 8] and color models [1], can be directly built into computation. For example, color models can be built into approximate computing when the result is encoded in color, such as heatmap. For another example, pixel-level perceptual models can be built into approximate computing where data is mapped to pixels, such as twitter map and some scientific visualization.

- Model user strategies when perceiving correlation in scatterplots

A visual feature was found to be an approximate proxy for judging correlation in scatterplots. However, one visual feature might not fully address human perception all the time. Thus, to accomplish perceptual models for correlation that widely scale requires the quantification of user strategies. The quantitative user strategies can have the power to predict perceptual models on a wide scale.

- Build a meta perceptual model for a visualization

  Perceptual models for correlation across multiple visualizations are successfully built and the vision science beyond was explored in my previous work. However, these models are limited to a specific task, which may constrict practical applications. A backward approach is to build meta perceptual models for all the tasks on one visualization, for example, a bar chart with two bars. This type of perceptual model is expected to be applied immediately in computation such as perception-based sampling.

- Measure the entropy of visualizations

  Some further and broad discussion of perception and design of visualization point to the concept of entropy, which could be used as a general measurement of how strong a pattern is. The entropy of visualizations may have the power to form a meta-explanation for perception of visualization and apply to computation for auto-binning.

- Build a theory-based animated transition system

  Beyond "data-visual" mapping, a more semantic and complete design model for visual forms could be raised from the standpoint of animated transition between visualizations. This model will be implemented as a system which enables any animated transitions between well-defined visualizations.

2 Personal Statement

2.1 My Background

I have a background and interest in art, engineering, human factors as well as degrees in Computer Science. The intersection of all of these naturally points to the fields of Visualization and Human-Computer Interaction, where I am right now.

My design background comes from my training of drawing for seven years, amateur design experience for five years and interests in the photography in recent years. My personal website posts some samples of my design work (hyperlink[1]).

My engineering background comes from my bachelor degree in engineering, where I built more than ten applications/systems such as online Othello game, PL/0 IDE and Student Information Manage System etc.

My human factors background comes from my early interests in psychology when I think understanding human is very interesting. I also selected some classes in psychology department at Tufts.

[1] https://www.eecs.tufts.edu/~fyang/index.html

My interdisciplinary background gives me the unique ability to think broadly and deeply and points me to the fields of Visualization and Human-Computer Interaction. My four years experience as a research assistant at Shandong University and Tufts University, confirms that I should stay in the fields.

2.2 Why Ph.D.

I grew up in an academic family. My father is a known professor of computational physics. My family members joke with each other by quoting physical laws.

From an early age, I have been a natural researcher. The questions I would ask my parents, teachers and professors, are more about "why" rather than "how". The way I think about the world, is first observing phenomena to find problems, then refining the problems, and solving them with the necessary help from others. I really like to find problems for myself and then solve them. Furthermore, the time when I was doing research at previous schools convinces me that I have the nature to do research.

A deeper reason for pursuing a Ph.D. is that I want to give something to the world during my lifetime. Although humans stay in the world for less than a century, technologies can stay definitely. I hope my scientific contribution could resist a longer time than my life so that I feel I'm still alive. I want my research to be known, to be recognized and to be used by the world. However, I know that my knowledge, background and resources are not strong enough to support me now. To obtain these, pursuing a Ph.D. is the only way for me.

2.3 Career Plan

During my Ph.D. time, first I will carry out my intended research and meet with the advisor's expectation and the department's requirements.

Beyond these, I would love to share my ideas with other people in the lab and out of the lab because I think ideas only get improved by taking other people's feedback. I also look forward to collaborating with people in the field and out of the field (i.e. perceptual psychology, cognitive science and machine learning), because more problems in the world (especially in visualization field) are interdisciplinary and can only be solved via cooperation between fields. In terms of research interests, I would love to join other projects that diverge from my topic but other people are passionate about, because I think knowing more and experiencing more are always good. More broadly, since I have a wide interest and background, I could provide help for other people in a variety of aspects.

After I graduate, depends on where I think I will be and the opportunities I will have, I may pursue post-doctoral fellowships or spend a few years in industry research labs. Alternatively, if I feel I am ready, I may directly go to academia and to be an assistant professor.

My ultimate career goal is to be a tenured faculty member in a research-oriented school and do self-motivated research with my students that have positive impacts to the world.

I have a dream that one day, I wake up on a warm beach with some breeze blowing my hair. When I open my eyes, someone there will greet to me: "Hey, nice to meet you Fumeng. You changed the world."

References

1. George A Agoston. *Color theory and its application in art and design*, volume 19. Springer, 2013.
2. Jacques Bertin. Semiology of graphics: diagrams, networks, maps. 1983.
3. Chun-Hsien Chou and Yun-Chin Li. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. *Circuits and Systems for Video Technology, IEEE Transactions on*, 5(6):467–476, 1995.
4. Lane Harrison, **Fumeng Yang**, Steven Franconeri, and Remco Chang. Ranking visualizations of correlation using weber's law. *IEEE Transactions on Visualization and Computer Graphics*, 2014.
5. Jock Mackinlay. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141, 1986.
6. Peter Shirley, Steve Marschner, Michael Ashikhmin, Michael Gleicher, Naty Hoffman, Garrett Johnson, Tamara Munzner, Erik Reinhard, Kelvin Sung, William B. Thompson, Peter Willemsen, and Brianl Wyvil. *Fundamentals of Computer Graphics*, chapter 27: Visualization, pages 675–707. A K Peters/CRC Press, third edition, Jul 2009.
7. **Fumeng Yang**, Lane Harrison, Ronald Rensink, Steven Franconeri, and Remco Chang. Why the percepion of correlation in scatterplots follows weber's law. *prepare for IEEE Transactions on Visualization and Computer Graphics*, 2016.
8. XK Yang, WS Ling, ZK Lu, Ee Ping Ong, and SS Yao. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication*, 20(7):662–680, 2005.