Lecture 15: Research Process and Paper Writing

Information Visualization
CPSC 533C, Fall 2011

Tamara Munzner

UBC Computer Science

Mon, 7 November 2011

Readings Covered

Process and Pitfalls in Writing Information Visualization Research Papers. Tamara Munzner. Chapter from Information Visualization: Human-Centered Issues and Perspectives. Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, Chris North, eds. Springer LNCS Volume 4950, p 134-153, 2008.

Reproducible Research in Signal Processing - What, why, and how. Patrick Vandewalle, Jelena Kovacevic and Martin Vetterli. IEEE Signal Processing Magazine, 26(3):37-47, May 2009.

Overview

- Writing InfoVis Papers: Pitfalls to Avoid
  - Pitfalls paper
- Non-Paper Research Process and Pitfalls
- Reproducable Research
  - Vandewalle paper
- Course-Specific Issues

Writing InfoVis Papers: Pitfalls to Avoid

- you should avoid them too!

Early Stage: Paper Types

- less useful for your final papers
- most course projects are design studies or algorithm/technique
- surveys, analysis not covered in this reading

Middle Stage: Visual Encoding

- Unjustified Visual Encoding
  - should justify why visual encoding design choices appropriate for problem
  - requires clear statement of problem and encoding, of course
- Hammer In Search Of Nail
  - characterize capabilities of new technique before submitting paper
  - even if start from technique-driven place
- 2D Good, 3D Better
  - must justify when benefits 3D outweigh cost of occlusion
  - abstract visual encoding allows choice over mapping variables to spatial position

Middle Stage: Visual Encoding 2

- Color Cacophony
  - blatant disregard for basic color perception facts
  - huge areas of highly saturated color
  - color coding intended for regions too small for distinguishability
  - nominal color coding for too many (15+) categories
  - red/green with no luminance difference
  - encode 3 separate variables with RGB
- Rainbows Just Like In The Sky
  - unjustified use of continuous rainbow colormap
  - hue does not have implicit perceptual ordering
  - standard rainbow colormap is perceptually nonlinear
  - for many nameable regions, quantize into segmented colormap

Later Pitfalls: Strategy

- What I Did Over My Summer Vacation
  - focus on effort not contribution
  - too low-level
- Least Publishable Unit
  - tiny increment beyond (your) previous work
  - bonus points: new name for old technique
- Dense As Plutonium
  - so much content that no room to explain why/what/how
  - fails reproducability test
- Bad Slice and Dice
  - two papers split up wrong
  - neither is standalone, yet both repeat

Later Pitfalls: Tactics

- Stealth Contributions
  - it's your job to tell reader explicitly
  - consider carefully, often different from original goals

Paper Writing: Contributions

- what are your research contributions?
  - what can we do that wasn't possible before?
  - how can we do something better than before?
  - what do we know that was unknown or unclear before?
- determines everything
  - from high-level message to which details
- often not obvious
  - diverged from original goals, in retrospect
- state them explicitly and clearly in introduction
  - don't hope that reviewer or reader will fill in for you
  - don't leave unsaid what should be obvious after close reading of previous work
    - pw very important - but many readers skip
  - goal is clarity, not overselling
    - do include limitations: often later, in discussion subsection

Later Pitfalls: Tactics

- Stealth Contributions
  - it's your job to tell reader explicitly
  - consider carefully, often different from original goals
- I Am So Unique
  - don't ignore previous work
  - both on similar problems and with similar solutions
- Enumeration Without Justification
  - "X did Y" not enough
  - must say why previous work doesn't solve your problem!
  - what limitations of theirs does your approach fix?
- Sweeping Assertions
  - cite source or delete assertion or flag as contrib
  - check what "everybody knows"
- I Am Utterly Perfect
  - discussion of limitations makes paper stronger

Later Pitfalls: Results

- Unfettered By Time
  - choose level of detail for performance numbers
  - detailed graphs for technique, high-level for design/eval
- Fear and Loathing of Complexity
  - present the complexity analysis for technique papers
  - full proof not required
- Straw Man Comparison
  - compare against state-of-the-art algorithms
  - head-to-head hardware best
- Tiny Toy Datasets
  - compare against state-of-the-art dataset sizes for technique
  - small datasets may be acceptable for user studies

Later Pitfalls: Results 2

- But My Friends Liked It
  - asking labmates not convincing when targets different
- Unjustified Tasks
  - user study tasks should be ecologically valid
  - convincing abstraction of real-world tasks of target users

Final Pitfalls: Style

- Deadly Detail Dump
  - how allowed only **after** what and why
- Story-Free Captions
  - optimize for flip-through-pictures skimming
- My Picture Speaks For Itself
  - explicitly walk them through images with discussion
- Grammar Is Optional
  - low-level flow is necessary (albeit not sufficient)
  - have native speaker check if you're ESL
- Mistakes Were Made
  - don't use passive voice
  - ambiguity about actor: your research contrib, or done by others?

Final Pitfalls: Style 2

- Jargon Attack
  - avoid where you can, define before using
- Nonspecific Use Of Large
  - hundreds, 10K, 100K, millions, billions?

Final Pitfalls: Submission

- Slimy Simultaneous Submission
  - often detected when same reviewer for both
  - instant dual rejection, multi-conference blacklist
- Resubmit Unchanged
  - often will get same reviewer, who will be irritated

Generality

- type: infovis
- encoding: color is general vis, others more infovis
- strategy: all research
- tactics: all research
- results: general vis
- style: all research, except
  - Story-Free Captions: general vis and graphics
  - My Picture Speaks For Itself: more infovis

Research Process and Pitfalls

- Review Reading
- Review Writing
- Conference Talks

Review Reading Pitfalls

- Reviewers Were Idiots
  - rare: insufficient background to judge worth
  - if reviewer didn't get point, many readers won't
  - rewrite so clearly that nobody can misunderstand
- Reviewers Were Threatened By My Brilliance
  - seldom: unduly harsh since intimately familiar area
- I Just Know Person X Wrote This Review
  - sometimes true, sometimes false
  - don't get fixated, try not to take it personally
- It's The Writing Not The Work
  - sometimes true: bad writing can doom good work
    - converse: good writing may save borderline work
  - sometimes false: weak work all too common
    - many people reinvent wheel
    - some people make worse wheels than previous ones

Review Writing Pitfalls

- Uncalibrated Dismay
  - remember you've mostly read the best of the best!
  - most new reviewers are overly harsh
- It's Been Done, Full Stop
  - you must say who did it in which paper
  - providing full citation is best
- You Didn't Cite Me
  - stop and think whether it's appropriate
  - be calm, not petulant
- You Didn't Channel Me
  - don't compare against the paper you would have written
  - review the paper they submitted

Conference Talk Pitfalls

- Results As Dessert
  - don't save til end as reward for the stalwart
  - showcase early to motivate
- A Thousand Words, No Pictures
  - aggressively replace words with illustrations
  - most slides should have a picture
- Full Coverage Or Bust
  - cannot fit all details from paper
  - talk as advertising, communicate big picture

Process Suggestions

- write and give talk first
- **then** create paper outline from talk
  - encourages concise explanations of critical ideas
  - avoids wordsmithing ratholes and digressions
- practice talk feedback session: at least 3x talk length
  - global comments, then slide by slide detailed discussion
  - nurture culture of internal critique
- have nonauthors read paper before submitting
  - internal review can catch many problems
  - ideally group feedback session as above

Paper Structure: General

- low level: necessary but not sufficient
  - correct grammar/spelling
  - sentence flow
- medium level: order of explanations
  - build up ideas
- high through low level:
  why/what before how
  - paper level
    - motivation: why should I care
    - overview: what did you do
    - details: how did you do it (algorithms)
  - section level
  - sometimes even subsection or paragraph

Reproducable Research

- 5. 15 minutes with free tools
- 4. 15 minutes with proprietary tools
- 3. considerable effort
- 2. extreme effort
- 1. cannot seem to be reproduced
- 0. cannot be reproduced

[Vandewalle, Kovacevic and Vetterli. Reproducible Research in Signal Processing - What, why, and how. IEEE Signal Processing Magazine, 26(3):37-47, May 2009.]

Why Bother With Reproducability?

- moral high ground:
  - for Science
- enlightened self-interest:
  - make your own life easier
  - you'll be cited more often

Levels To Consider

- paper
  - post it online
  - makes sure it stays accessible
- algorithm
  - documented in paper itself
  - document further with supplemental materials
- code
  - make available as open source
- data
  - make available
  - vis tricky issue: data might not be yours to release!
- parameters
  - how exactly to regenerate/produce figures, tables

Course-Specific Issues

Updates

- alphabetical by first name, Nov 14/16/21
  - Mon Nov 14
    - Anna, Anton, Chuan, Jesssica
  - Wed Nov 16
    - Jillian, Jingxian/Junhao, Joel, Louise
  - Mon Nov 21
    - Mashid, Niels, Shama
- four per day
  - 18 minutes each: 15 min talk, 3 min questions
  - (end 20 min early on third day)
- by 11am send email
  - either with your slides,
  - or telling me you're using your own laptop
    - in that case, slides to me by 6pm that day

Individual Meetings

- I encourage you to meet with me before final presentation
  - chance to get feedback when you can still act on it!
  - optional, not mandatory
- particularly good times
  - partway done, several weeks after updates
  - mostly done, week or so before due
- schedule ahead by email (best), or use office hours

Final Presentations

- context
  - department will be invited
  - refreshments will be served
  - order: alphabetical by first name
- 15 min: 12 minutes talk, 3 minutes questions
  - some context setting, but focus on results
  - ok to assume audience already saw update
- demos encouraged
  - do include screenshots in slides as backup
  - practice timing in advance since hard to do quickly
  - if you're using my laptop, must do checkout in advance

Final Project Writeups

- no length restrictions
  - use images liberally
- conference paper format
  - use templates provided (LaTeX, Word)
  - submit PDF
- due two days after presentations (Wed 12/14 noon)
- standalone document
  - ok to reuse some text from proposal (only if appropriate)
- please do read Project Description page closely!

Final Project Writeup Structure

- Introduction - description of problem: task, data
- Related work
- Description of solution: infovis techniques, visual encoding
- Medium-level implementation
  - must include specifics of what other components or libraries you built upon, vs. what you did yourself
- Results
- Screenshots of your software in action
- Scenarios of use
- Discussion and Future Work
  - strengths and weaknesses
  - lessons learned
  - what would you do if you had more time?
- Bibliography

Course Requirements vs. Standard: 1

- research novelty not required
  - some past projects implement published technique
  - some past projects explicitly not aiming for academic publishability
  - many past projects propose solution using existing techniques (design study)
  - some past projects extend/refine algorithms (technique)
  - some past projects have become posters at InfoVis
  - some past projects could have been submitted as papers with further work

Course Requirements vs. Standard: 2

- explicit explanation of what was coded is required for programming projects
  - submission of code is also required
  - you're encouraged but not required to make project available open-source!
- part of my judgement is about how much work you did
  - high level: what toolkits etc did you use
  - medium level: what pre-existing features did you use
  - medium level: how did you adapt/extend existing features to solve your specific problems
- design justification is required (unless analysis/survey project)
  - technique explanation alone is not enough

Course Requirements vs. Standard: 3

- user studies not required - time frame too short
- confirm that your color choices appropriate
  - vischeck.com for colorblind
  - legibility, color guidelines

Writing Correctness and Style

- http://www.cs.ubc.ca/ tmm/writing.txt

Code

- pack up with tar/gzip/zip
- must have top-level README with roadmap for files
  - which parts are your code, which are libraries, etc
  - how to compile
  - how to run
- acceptable that it doesn't compile on my machine if you targeted another platform