



# Reading Scientific Papers

Scientific Soft Skill Seminar

Petr Kmoch

CGG MFF UK

# Reading Scientific Papers

- **What?!?**
- **It's simple, right?**

“Read them,” said the King.

“Where shall I begin, please your Majesty?”

“Begin at the beginning,” the King said gravely, “and go on till you come to the end, then stop.”

—Lewis Carroll, Alice in Wonderland

- **Well, not quite ;-)**

# Presentation Outline

---

- Reading goals & types
  - With examples
- Paper processing
- Useful tips

# Presentation Outline

---

- **Reading goals & types**
  - **With examples**
- Paper processing
- Useful tips

# Typical Paper Structure

---

- Abstract
- Introduction
  - Topic overview
  - Contribution summary
- State of the art
- Contribution
- Results
- Conclusion
- Future work

# Goals of Reading

---

- Learn specific info
- Keep up to date
- Assert novelty
- Broaden perspectives
- Write a review
- Implement

# Reading Types

---

- Reading types
  1. Scan (what)
  2. Read (how)
  3. Save
  4. Learn (why)
- Different purpose  $\approx$  different type

# 1. Scan

---

- Goal: Evaluate paper relevance
  - Output: now / later / not at all
- Title, authors, abstract, introduction, results, conclusion
- Year
- Section titles, figures
- Trust authors' claims

# 1. Scan – Practical Tips

---

- Title “keywords”
  - “implementation”, “application”
    - existing method without changes
  - “survey”, “overview”, “review”
    - just STAR
  - “X-based Y”
    - X is old, Y can be new
  - “framework”
    - multiple method compilation, down-to-earth results
    - often many self-citations

## 2. Read

---

- Goal: Understand what is done & how
- More than one reading necessary
- Connection to other papers
- Formulae, descriptions
- Be doubtful about authors' claims
  - Look for “fine print”
  - Find weak spots
    - Look in Future work

## 2. Read – Practical Tips

---

- Parameter values not given
  - difficult to tune; far from real-world values
- Important parts missing
  - referred to other papers
  - referred to “textbooks, literature, ...”
- No implementation
  - look for concrete results (timings, precision, ...)
- Narrow comparison range
  - method is data-sensitive
- Text instead of formulae
  - formulae complex, illogical or ad hoc
- Claims rebutted by newer papers

## 3. Save

---

- Goal: Create reference for the future
- Summarize paper in your own words
- Annotate important formulae
- Stress pros & cons
- Write down full BibTeX entry
  - Rendered bibliographic entry also useful

# 3. Save – Practical Tips

---

- Later...

## 4. Learn

---

- Goal: Understand the paper inside out
- Don't believe any authors' claims
  - Unless you can prove them
- Cross-reference formulae
- Write down full notation explanation
- Read (type 2) all papers detailing methods employed

## 4. Learn – Practical Tips

---

- Table of symbols
  - Know the **why** of formulae
- Implementer's view
  - Data flow
  - Effects of “sketched” steps
  - External dependencies
- Resource-intensity
  - Evaluate combinability

# Presentation Outline

---

- Reading goals & types
  - With examples
- **Paper processing**
- Useful tips

# Paper Sources

---

- Faculty library
- Digital libraries (ACM, IEEE, EG)
  - Faculty has (some) access
- State technical library, CAS library
- CiteSeer(X)
- Authors' webpages
  - Not always the first author's
- Conference websites

# Number Explosion

- Easily 1000s of papers
- Processing ways
  - Backtrack references
  - Method – old vs. Survey – new
  - Reference age
  - Venue, author affiliation

## List of Citations from Science Citation Index for

M. Kass, A. Witkin, and D. Terzopoulos, "Snakes - Active Contour Models" International Journal of Computer Vision, 1(4): 321-331, 1987.

1988: 2 1989: 1 1990: 1 1993: 11 1994: 25 1995: 37 1996: 74 1997: 99 1998: 88  
1999: 115 2000: 129 2001: 129 2002: 134 2003: 172 2004: 194 2005: 27

**Total citations:** 1238

As of 11 Mar 2005

[http://iacl.ece.jhu.edu/projects/gvf/gvf\\_cite/snake\\_cite\\_year.html](http://iacl.ece.jhu.edu/projects/gvf/gvf_cite/snake_cite_year.html)

# Paper Management

---

- Organizing saved data (type 3)
- 100s can accumulate fast
- Stand-alone database
  - Files, actual DB, ...
- BibTeX database
  - Custom fields
  - Manager SW: JabRef, ...
- Hard copies with hand scribbles

# BibTeX

- Structured text file
- Entries
  - Article, Book, InProceedings, PhDThesis, ...
- Fields
  - title, author, year, publisher, ...
- Natively cooperates with LaTeX
  - Different styles
- MS Word integration exists

# Sidestepping Reading

---

- Knowledge needed
  - Find a citing STAR or book
  - Find an “X-based” paper
- Implementation needed
  - Authors’ webpages
  - Libraries where authors contributed
  - Software docs can include citations
  - MATLAB, *Mathematica* prototyping

# Presentation Outline

---

- Reading goals & types
  - With examples
- Paper processing
- **Useful tips**

# Reading Approaches

---

- Papers vs. books
  - Cutting edge vs. tried & tested
- Conference, journal, invited, STAR, thesis
- User's explanation
  - Method explained better in a citing paper
- Semi-unrelated fields
  - Knowledge transfer & novel application

# Keyword Dictionary

---

- Real-time
  - $\geq 20$ -25 FPS
- Interactive
  - 5-10 FPS
- Semi-automatic
  - (Expert) user cooperation necessary
- Out of core
  - Too large for main memory

# Keyword Dictionary (cont.)

---

- Image-based
  - 2D, post-processing, IP methods
- Sketch-based
  - User-interactive, fuzzy
- Multi-dimensional vs. High-dimensional
  - 4-5 vs.  $O(10^2)$
- Approximation
  - Expect ad hoc formulae and “magic numbers”

# Keyword Dictionary (cont.2)

---

- Analytical, closed-form
  - Expressed as a formula
- Pre-computed
  - Intensive pre-processing phase
- Algorithm vs. Solution
  - Solution can be theory-only
- Parallel vs. Distributed
  - Intra-case vs. Net-based

# Thank you

---

- Questions?
- Answers