1IMV00 Seminar Visualization – Study guide 2016/2017

1.1 General
This document gives information on the course 2IMV00 Seminar Visualization. It will be given in the second quartile of the 2016/2017-course year. It is targeted at second year students of the Master programs Computer Science & Engineering and Business Information Systems of TU Eindhoven, and gives 5 ECTS credits.

1.2 Aims
In this seminar, students will get acquainted with research in the area of visualization. Papers and/or book chapters will be studied and presented by students. As a result, students will obtain knowledge on some recent developments in the areas described. More important however, is that students have gained experience in finding, reading, reviewing, and digesting literature; as well as writing and reviewing research papers.

1.3 Lecturer
Prof.dr.ir. Jack van Wijk, MF 4.059, tel: (040) 247 4579, e-mail: j.j.v.wijk@tue.nl

Feel free to contact me if questions arise. Safest is to make an appointment via email, but just checking if I am in my office and have time is fine as well.

1.4 Format
The format given here is based on the assumption that a limited number of students will participate. If more than 12 students want to follow the seminar, changes might be needed. Each student will individually study one particular topic, write a survey article, and give a presentation.

1.5 Schedule
Below a tentative schedule is given. Meetings will be held on

- Wednesdays, 7th and 8th hour (15:45-17:30), Helix Oost 1.91, and on
- Fridays, 3rd and 4th hour (10:45-12:30), Helix Oost 2.91.

These meetings will not all be held, see the schedule below. Students are expected to attend all planned meetings. If this is not possible, send a message (with a good reason) in advance. If custom advice is required, feel free to contact the lecturer. The deliverables mentioned are described in more detail in following sections. Students should stick to all deadlines mentioned. Especially the last one, for the final version, will be strictly maintained.

Week 1
Wednesday 16 November Introduction
An overview of the course is given.

Students study the list of topics, and prepare a motivated choice of the topic they want to investigate, as well as a second and third choice. We aim at all students studying different topics, so take this second and third choice seriously.
Week 2
Students submit at the latest **Tuesday 22 November 12:00** via OASE a proposal for their choice of topic.

**Wednesday 23 November Progress meeting**
Topic choices are discussed, challenges and pitfalls are identified, and possible conflicts are resolved. Progress is discussed, experiences are exchanged, and advice is given.

Week 3
Students submit at the latest **Tuesday 29 November 12:00** via OASE an outline of their survey article.

**Wednesday 30 November First short presentation**
The next day these are discussed. Students prepare a small five minute presentation, with a few slides, to present what they found and to sketch the big picture. Also, the schedule for the final presentations is filled in.

Week 4
**Wednesday 7 December First short presentation (cntd.) / progress meeting**
If needed, presentations are continued, to be decided on the November 30th. Progress is discussed, experiences are exchanged, and advice is given.

Week 5
**Wednesday 14 December Progress meeting**
Progress is discussed, experiences are exchanged, and advice is given. Recommendations for the final presentations are given.

Students submit at the latest **Friday 16 December 12:00** via OASE the first complete version of their survey article.

Week 6
**Wednesday 21 December Discussion first versions**
Feedback on first versions is given and discussed with students.

In the weeks after the winter break, students present their findings. Schedule to be announced.

Week 7
**Wednesday 11 January Presentations**
**Friday 13 January Presentations**

Week 8
**Wednesday 18 January Presentations**
**Friday 20 January Presentations**

Students submit at the latest **Friday 3 February, 12:00** (two weeks after the last classes of quartile 2) via OASE the final version of their survey article.
1.6 Grading
Students will be evaluated on the final version of the survey article (60%), the presentation (30%), and their attendance and participation during meetings, especially concerning active listening to the presentations (10%).

2 Deliverables
A general request for all deliverables: Make sure that each contains your name, e-mail address, student number, and a date. Submit your documents as pdf or as Microsoft Word documents via OASE.

2.1 Topic choice
This document should contain:

- The title of your topic;
- A research question;
- A short description, including the background of the problem (such as typical applications), reasons why this is an interesting issue, and resources (conferences, research communities) you want to explore.
- One or more references to relevant articles.
- The title of a second and third choice of topic, as an alternative when there are issues with the first one, especially multiple students that want to do the same topic.

In section 3 a number of suggestions for suitable topics in Information Visualization and Scientific Visualization are given. Topics from related areas (computer graphics, animation, user interfaces) are also possible, for instance when there is a strong personal interest or if one wants to use this seminar as a start-up of a final project.

Maybe the hardest problem in defining a topic is to find one with a proper size. A topic like “Information Visualization”, or “Medical Visualization” is far too broad. Hundreds of articles have been written about this, and only a shallow overview can be provided in the time available. Students are encouraged to select a much narrower and focused topic, such that searching and finding literature becomes a challenge on its own, and that different approaches and solutions can be compared at a detailed, technical level. Think in advance how the topic could be expanded or more focussed, and write down your ideas in the initial document. If the initial choice turns out not to be optimal (for instance, too broad or too narrow), this can be discussed at one of the meetings or in between. But on the other hand, don’t give up too quickly.

Also, check if you can find overview articles that have been written already. If there is one, recent, solid overview – select another topic. If the overview is somewhat older (say, before 2005), check if there are new developments, such that you can provide an extension of this overview. If there are several overviews, you can try to write a meta-overview, comparing the distinctions made in different articles. But take care, this is not easy, as it requires you to read not only the overview articles, but also the papers these refer to.
2.2 Outline survey article

This document should describe the framework of the article:

- Title
- Abstract
- Titles of sections and sub-sections, with a short, bullet-wise description of the content
- Schematic diagram of the structure of the field
- References to literature read
- Enumeration of loose ends and plans for further exploration

This document serves several purposes. It aims at providing a stimulus to start working in time, and as a basis for discussion of the progress made. By the time this document is produced, the student should have scanned about 40-60% of the relevant papers, and have a clear vision how to structure the material obtained from the literature. Bring a stack of collected papers with you when these outlines are discussed. If you think it is useful to expand or reduce the original topic, give an argumentation for that.

To encourage students to take this seriously and to start in time, a short presentation is expected, where the student gives an overall view on his topic and what he intends to write. Five minutes and a few slides should do. Important: Use diagrams, schemes, tables etc. to visualize how you view your topic.

2.3 Article

The overall aim is to produce something that looks, smells, feels like and maybe even is a scientific article. The following holds for the first, as well as the final version of the survey article. The first version should be complete, the next round should be used to address issues raised.

2.3.1 Format

The article should be 8-10 pages long, pdf, in IEEE VIS format. Formatting guidelines can be found at [http://www.cs.sfu.ca/~vis/Tasks/camera_tvcg.html](http://www.cs.sfu.ca/~vis/Tasks/camera_tvcg.html). This is the format used for the IEEE Visualization and InfoVis conferences, and is very close to the format of IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG). These are the leading conferences and the leading journal in visualization. Use of this format will lead to a result that is visually comparable to many of the papers you will read. Eight to ten pages do not seem many, but you will find out how much information can fit on double column pages with 9 point text. Take care of figures and illustrations, don’t use too few or too many. A tip on printing papers in this format: Use “no scaling”. The format is for the US letter format, but the contents fit when printed without scaling on A4.

2.3.2 Scope

A survey can be done for two reasons. Firstly, one can aim to solve a problem. For this, a good understanding of the problem is needed, as well as an overview of existing solutions. The article should answer the question which solution is most appropriate given certain requirements and constraints (type of data, scale, effort needed, quality of result, etc.). Secondly, if one has the ambition to do research in
an area, it is important to understand what has been done and what opportunities exist for improvement.

Students are encouraged to address both issues: Make clear what has been done and what is useful, given circumstances, and also, give recommendations for future research. For the first, one can think about a more practical project to be done in industry; for the second, consider a research project at the university. What final masters project would you formulate?

2.3.3 Structure
Per discipline, articles have a standard structure. A conventional Visualization paper, describing a method, tool, or technique has the following structure:

1. Introduction (problem, application areas, overview)
2. Background (issues, requirements, existing approaches)
3. Concept (approach used here)
4. Details (elaboration of the approach)
5. Results (results and comparison with existing approaches)
6. Conclusions (summary of findings, suggestions for future work)

Obviously, the fourth section often contains the main content, and can be split into multiple sections. One can deviate from this overall structure, but then reviewers will complain. Survey papers are relatively rare. A possible structure is:

1. Introduction (problem, application areas, overview)
2. Background (issues, requirements, existing surveys)
3. Analysis (Segmentation of the problem: what kind of data, what visualization choices?)
4. Details (Discussion how current approaches fill in various aspects)
5. Conclusions (summary of findings, for instance in a table form, suggestions for future work)

It is tempting just to give an enumeration of summaries of papers. This should be avoided! This does not lead to a good survey paper. Rather, the analysis counts (what is a natural and useful subdivision, taxonomy or structure), and next the topic is addressed by a discussion per aspect. This is often not easy, but this is crucial to really understand the structure of the problem as well as the existing solutions. For inspiration, examples of solid (and long) recent overview papers are:


EuroVis has a special track for State of the Art reports, where more examples can be found.
2.3.4 **Text**
Text should be original. Only if absolutely necessary, short, literal quotations with proper reference are acceptable. *Copy-paste of paragraphs of text without reference is considered as fraud*, and will be handled appropriately.

2.3.5 **Images**
Do not use not too many, and also do not use too few images. Do not use images as page-filling, without being referenced or discussed in the text. Give proper references to images. Use schemes and diagrams to show the structure of the problem, draw these yourself when needed. Make sure all text in images is legible. Avoid the use of raster images (.jpg) for schematic figures, use a vector format (.ps, .pdf) instead.

2.3.6 **Reading**
Critical reading is important. Although articles that have appeared in journals and major conferences are peer-reviewed, they still can contain flaws and errors. Also, later results might show that earlier results were limited, and, the same topics can have been studied in different communities. Carefully check if statements made are valid and true. This holds especially for the quality of the results obtained. If the authors claim that some approach provides new insight or is easy to use: Did they verify this with a user study? Has this user study been properly set up? If performance results are presented, are the data sets used representative? When authors describe previous work: Is this done correctly and fair? One way to verify the material presented in the literature is to implement methods yourself, and do a comparison. For many topics this will require too much effort, but in some cases this can done in reasonable time and will provide much insight. Also, the world is much larger than the academic world. If a technique from 1990 claims to be superior – has it been picked up in the real world? If not, why? Also, outside the research community also new methods are developed. Such work is usually not formally reviewed and presented in a standardized way, so be critical here also.

2.3.7 **Sources**
Solid books on Information Visualization, with – rightfully – much attention to perception issues, are:


Various websites give overviews of useful sources and key-publications:

http://www.infovis-wiki.net
http://www2.sims.berkeley.edu/courses/is247/f05/resources.html
http://www.cs.umbc.edu/~rheingan/636/readings.html
http://www.cc.gatech.edu/~john.stasko/7450/biblio.html
http://www.visualcomplexity.com
For visualization, the key conferences are IEEE Visualization, IEEE InfoVis, EG EuroVis, PacificVis, and IEEE VAST, the key journal is IEEE TVCG. Papers from these can be retrieved the Digital Library of IEEE, via the TU/e library site. But, many other conferences also accept papers on visualization, for instance the Graph Drawing conference, ACM CHI, ACM UIST, and ACM SIGGRAPH.

To get a quick overview of papers written by a certain author, you can use DBLP: http://www.informatik.uni-trier.de/~ley/db, another powerful search tool is Google Scholar (http://scholar.google.com). This site allows you to do backward tracing (which papers are cited) as well as forward tracing (which papers have cited a paper). Also, consult web-pages of authors, they often contain recent work and material. And of course, Google itself is highly useful for literature search. Make sure that you focus on reviewed publications, and refer to those publications, and not to the web pages.

And finally, besides in the scientific literature, also much information can be found outside, in blogs, at web-sites, in reports, etc. For instance, the New York Times produces many high quality infographics, which can be inspiring; people like Stephen Few, Robert Kosara, and Andrew Vandemoere are active bloggers on information visualization.

2.4 Presentation
The presentation should take about 20 minutes, followed by 5 minutes for technical questions. A beamer is available. Next, the presentation itself is discussed for 15 minutes. Students are expected to listen critically, and bring forward stronger and weaker aspects of the presentation.

3 Suggestions for Topics
The field of visualization can be subdivided in various ways, for instance:

- Per application. Different approaches are pursued dependent on the problem to be solved by the end-user, which leads to subdisciplines like chemical, medical, geographic visualization;

- Per data type. For many different types of data types dedicated visualization methods have been developed. This gives subdisciplines like multi-variate visualization (for tables), tree visualization, volume visualization, and vector field visualization. Special challenges here are combinations of data types, and handling of dynamic data.

- Per mapping method. A large number of standard approaches are available to map data to geometric and graphic objects, such as scatterplots, node-link diagrams, streamline plots and iso-surface displays. Many variations and refinements on all these have been developed.

- Generic topics. Some issues, like evaluation and taxonomies, concern all types of visualization.

Below a number of possible topics are suggested, divided over these categories. Per topic the title, a short description, and a single reference is given as a starting point. See further the remarks given in section 2.1.
3.1 Application

Web usage. How can visualization of logged web usage data help to get insight and to improve web sites? How to understand how people navigate?

Cohort selection. Selecting groups of patients with similar characteristics and comparing these is standard practice in medical research. But how to deal with messy data, like images, and how to prevent bias?

Databases. Databases often contain hundredths of tables, and thousands of columns. How can we get global insight in the sort of data that is in the database?

Twitter. Twitter is an interesting source of information about what people are interested in. How to get insight in this?

Financial transactions. Such data sets are huge, and a clever combination of automated and visual methods is needed to get more insight in patterns and fraud.

News. Take some recent event that appeared in the news (a sports tournament, earthquakes, refugees, financial crisis, demographic developments), and study how these can be visualized. Probably not much scientific literature can be found, a critical mind and eye is needed.

3.2 Data Types

Uncertainty visualization. Real-world data is often uncertain and not precise. Many different solutions have been proposed for their visualization, but is this solved now? Several overviews have appeared, select a subtopic and go in depth. G.-P. Bonneau, H.-C. Hege, C.R. Johnson, M.M. Oliveira, K. Potter, P. Rheingans, T. Schultz. Overview and State-of-the-Art of Uncertainty Visualization. Chapter 1 in C.D. Hansen et al. (eds.) Scientific Visualization - Uncertainty, Multifield, Biomedical, and Scalable Visualization. Springer, 2014.
**Time series.** Time series are ubiquitous. They can be visualized with standard line graphs, but when such datasets are large, complimentary methods are needed to visualize and interact with them.


**Video visualization.** How to present a video in a compact representation, such that features stand out?


**Association rules.** A classic application of data mining is to find association rules: If customers buy wine and olives, they probably also buy cheese. Thousands of these can be found, but how to support the user in picking relevant ones?


**Neural networks.** Another classic in data mining. Can we understand what’s going on? Also when they are not used for image classification?


**Linear Regression.** A relatively simple model for analysing data is linear regression, but many variables and showing the underlying data properly can be difficult.


**Word counts.** Given one or more collections of documents, how to show which words are most frequently used? The Wordcloud is one solution, but there are alternatives.


**Clusters.** Clustering is a standard approach to analyse and understand multivariate data. However, many choices have to be made. How can visualization be used to support this? How to show clusters? Jinwook Seo, Ben Shneiderman. Interactively Exploring Hierarchical Clustering Results. IEEE Computer 35(7), p. 80-86, 2002.

**Network comparison.** Given two networks, how to show their differences?

3.3 Mapping

*Parallel coordinate plots.* These are a simple, but generic approach to visualize multi-dimensional data. However, they suffer from poor scalability. What can be done to remedy this? The reference below gives a solid overview, select some subtopic and go in depth.


*Large graphs/performance.* Recently, much progress has been made in accelerating algorithms for drawing large graphs. How do these work? And what is the price to pay?


*Matrix representations of graphs.* One way to visualize graphs is to use incidence matrices. What variations have been developed, what are the pro’s and cons?


*Event visualization.* Given a sequence of events, each with a time stamp and attributes. Translate them into icons, plot them on a timeline, and you’re done. Or is there more?


*Orderability.* Different cues can be used to communicate data, but not all of them are equally useful. If the data is numerical or ordered (small-medium-large), we should use cues that are perceptually easy to order. What to use?


*Sensitivity.* Different cues for visualization have different sensitivity, especially on the number of steps that can be distinguished easily. What is known about this? Much research has been done in psychophysics, the topic is huge. Aim at a broad overview, go in depth for a few cues (intensity, hue), or focus on experimental methods to study this.


*Physical Data Visualization.* Besides as images, data can also be visualized with physical objects. What has been done in this area, and what are possible advantages?

3.4 Generic

*Evaluation.* The aim of visualization is to obtain insight. But what is insight, and how to measure this in a controlled setting? Large topic, specialize.


*Tasks.* What are typical tasks in visualization, and how can we sort these out? Different authors have different opinions on this.


*Storytelling.* Infographics tell a story about data. What patterns can we distinguish? The paper below gives a overview, is this complete and do you agree?


*Visual literacy.* The success of a visualization depends on the skills people have to interpret these, known as visual literacy. But, this fuzzy term is hard to pin down. What have people in the infovis community to say about this? How to assess it?