EG2013 Tutorial on VIDEO VISUALIZATION

6. Empirical Studies and User Evaluation

Kuno Kurzhals

University of Stuttgart

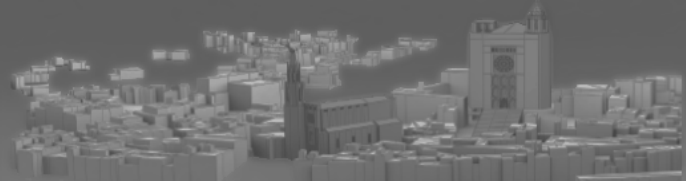
Why Evaluating Video Visualization Techniques?

- Understand the fundamentals
  - How does the visualization convey information?
  - When/where does it become more efficient or effective than other technologies (e.g., data mining, computer vision)?

- Gain insights
  - What makes one type of visualization work better than others?
  - What design guidlines can be inferred from an empirical study?

- Usability evaluation
  - Does a visualization improve users' ability to perform a task?
  - Can users learn to use a type of visualization?

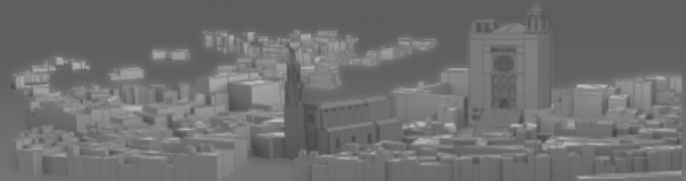
Challenges of Gaining Insight



- What comes first?
  - Asking the right questions
  - Finding the right method
  - Choosing the right task

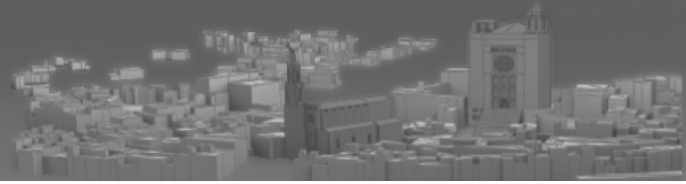
- Desirable factors in a study[Carpendale08]:
  - **Generalizability**: Result applies to other people/situations
  - **Precision**: Precise measurements, control of confounding factors
  - **Realism**: Study and use case are in the same context

Important Steps[Forsell10]



- Experimental Design
  - Between-subject, within-subject or mixed
- Tasks
  - Generalizability
  - Choice of stimuli (real or synthetic)
- Participants
  - Number of participants
  - Target group
- Assignments
  - Order of presentation
  - Random, counterbalance
- Results
  - Descriptive & inferential statistics

Statistics



- Typical questions
  - Is the result from my visualization different than from another one? (better?)
  - Are there correlations between variables?
- Different variable types
  - Independent (e.g. different visualizations)
  - Dependent (e.g. completion time for a task)
- Statistics
  - Descriptive: e.g. mean, median, standard deviation
  - Inferential: Statistical tests to find significances

Summary of Different Evaluation Methods

Methods and examples in this tutorial:

- Quantitative, Controlled Lab Studies
  - Task performance measure
- Quantitative and Qualitative Evaluation with Questionnaires
- Case Studies and Expert Evaluation
  - Think aloud
- Eye-Tracking with Video Stimuli

Quantitative, Controlled Lab Studies

Quantitative, Controlled Lab Studies

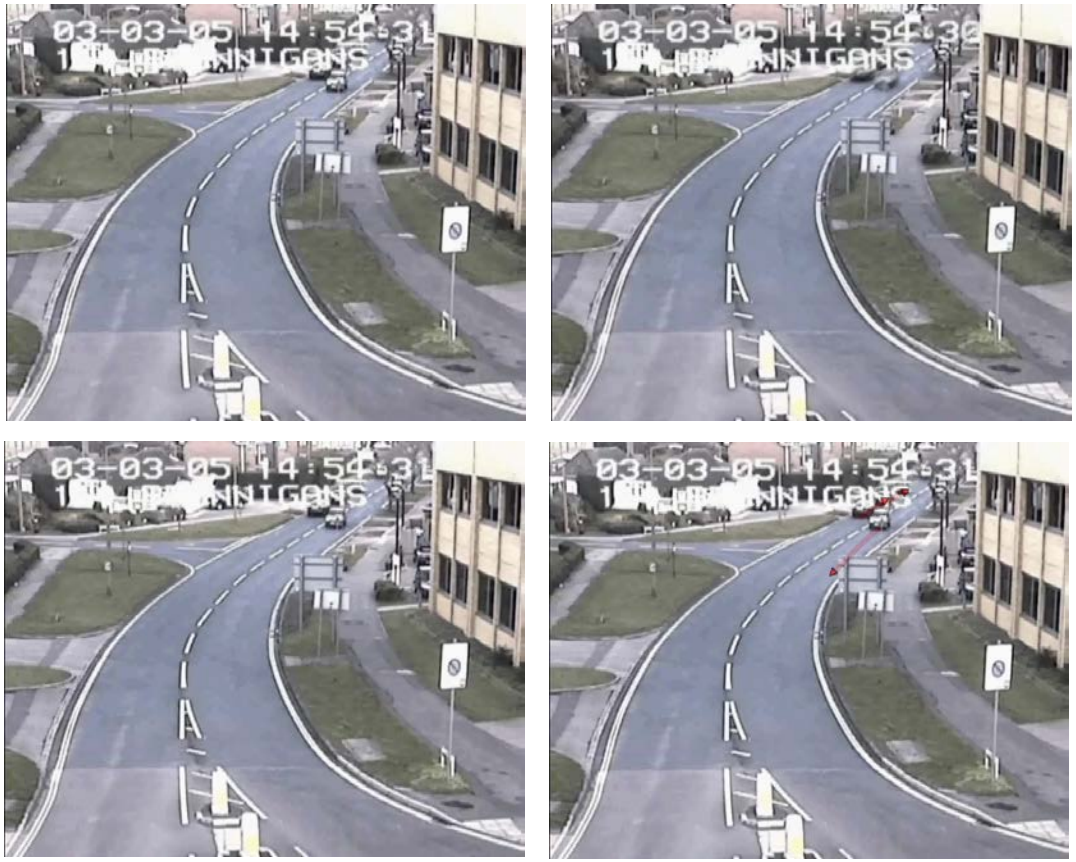
Quantitative, Controlled Lab Studies

Task Performance Measure

- User study under controlled, restrictive conditions
- Reliable results, but often focused on one aspect
- Design of task
  - Confounding factors
  - Leads to effects on dependent variables
  - Example: Search & report
- Measure and protocol
  - Accuracy : Correct and false answers
  - Efficiency : Completion times, reaction times

Example: Fast-Forward Video Visualization [Höferlin12]

- Visualizations to improve video fast-forward
- Question:
  - What influence have the visualizations on object identification?



[Höferlin12]

Example: Fast-Forward Video Visualization [Höferlin12]

- Task
  - Cartoon character appears several times in the videos
  - Participants use a buzzer to confirm identification

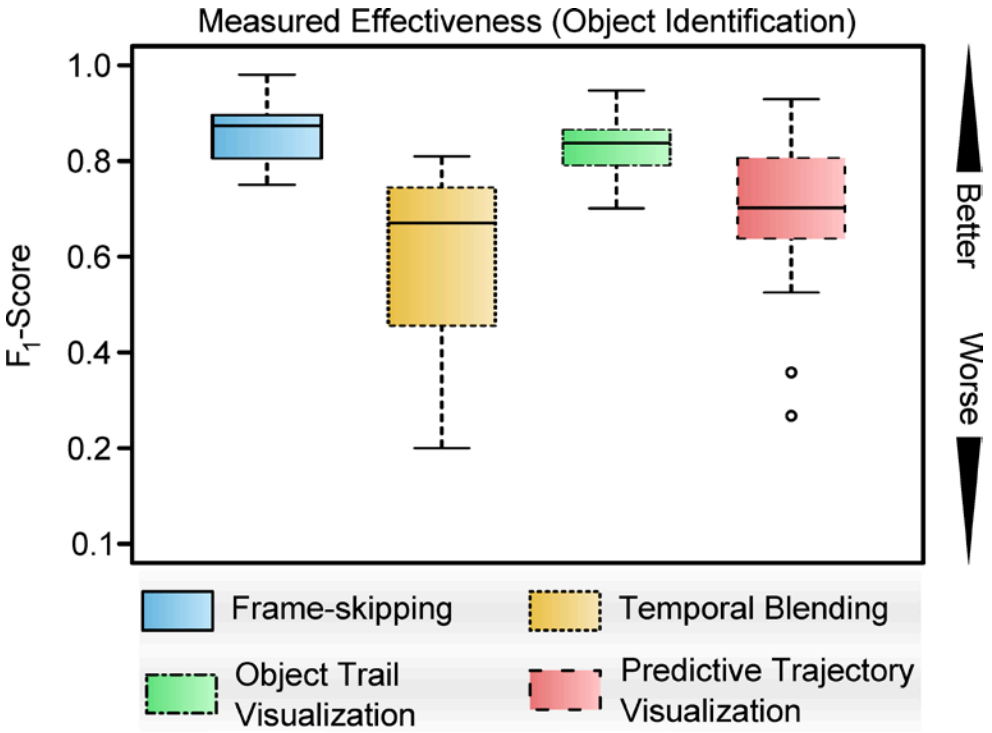
[Höferlin12]

Example: Fast-Forward Video Visualization[Höferlin12]

- Pilot study: 4 participants
  - Task difficulty improved:
    More cartoon characters & faster videos
- User study: 24 participants
- Within-subjects design
  - All 4 visualizations per participants
- Counterbalancing
  - Visualization order & video stimuli randomized
- Measures
  - Accuracy: Buzzer logs
  - Comfort & preference: Questionnaire
  - Efficiency measure possible

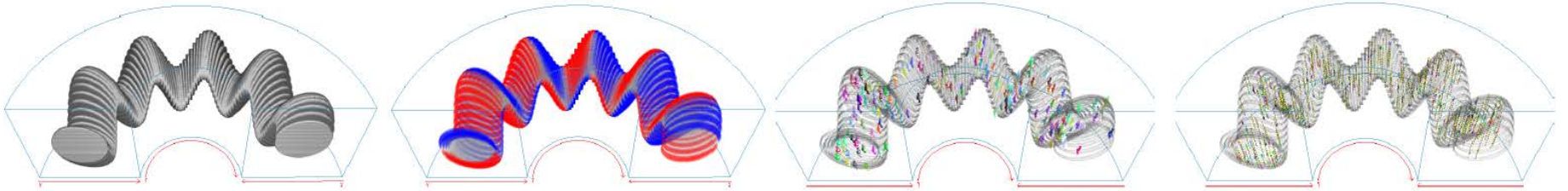
Results

- Statistical testing
  - Dependent variable: Accuracy score
- Choosing a test
  - Normal distribution: ANOVA
  - Non-parametric tests (e.g. Kruskal-Wallis)
- Post-hoc testing
- Significant effect of visualization on scores
  - Conventional method best scores
  - Blending worst score



[Höferlin12]

Example: Video Signatures[Chen06]

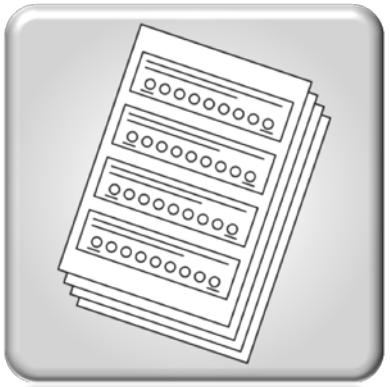


(a) Type A: silhouette hull (b) Type B: 4-band difference volume (c) Type C: motion flow with glyphs (d) Type D: motion flow with streamlines

- 69 participants
- Main objectives:
  - Can users learn to recognize motions from their visual signatures?
  - Obtain data that measures difficulties and time requirements of a learning process.
  - Evaluate the effectiveness of four types of visual signatures.
- Procedure:
  - Sessions with oral presentation and examples
  - Indentify underlying motion patterns (speed & accuracy)
  - Choose visual signature with most relevant information for different motion clips
  - Supplementary user study

Quantitative and Qualitative Evaluation with Questionnaires

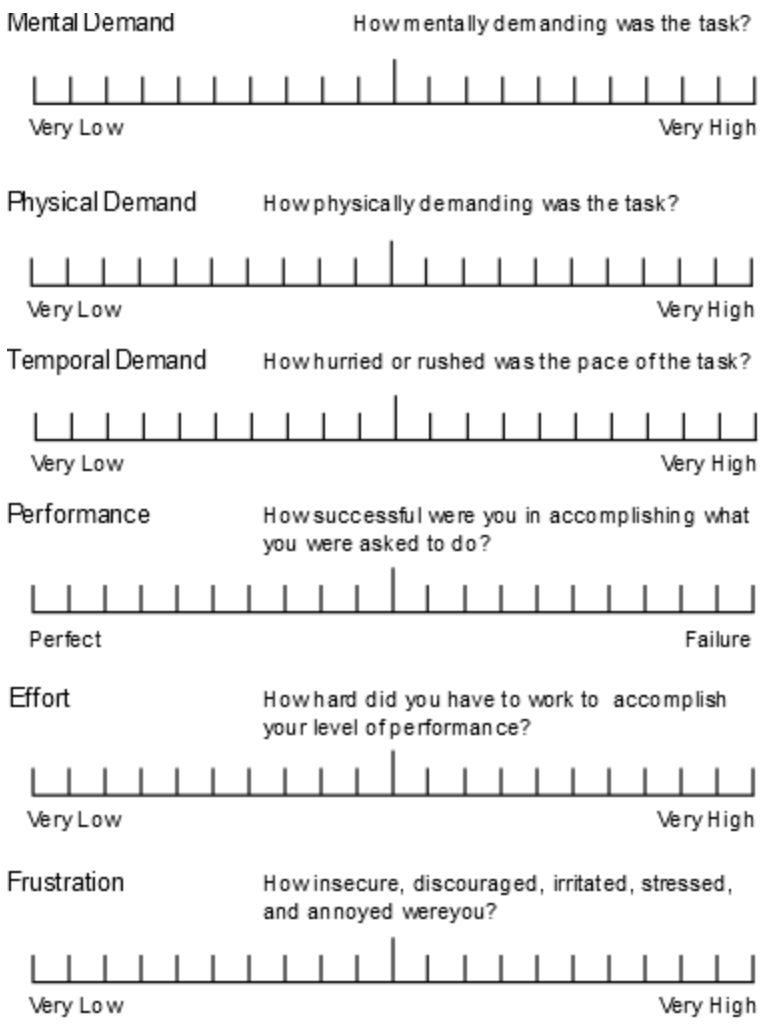
Why use Questionnaires?



- Not everything is objectively measurable
  - Frustration
  - Mental demand
- Evaluation of subjective impressions
  - Qualitative: Open-ended questions
  - Quantitative: Likert scales
- Example: NASA TLX

Example: NASA TLX[Hart88]

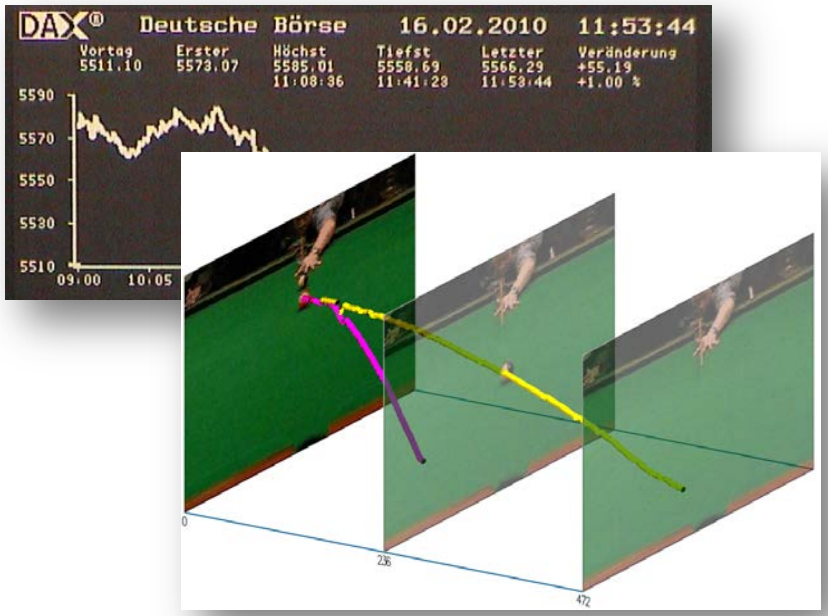
- Task Load Index
- 6 generic, task-related factors
- Biploar scales from very low/very high
  - Exception: Performance (perfect/failure)
  - 20 equal intervals
- Weighting of factors

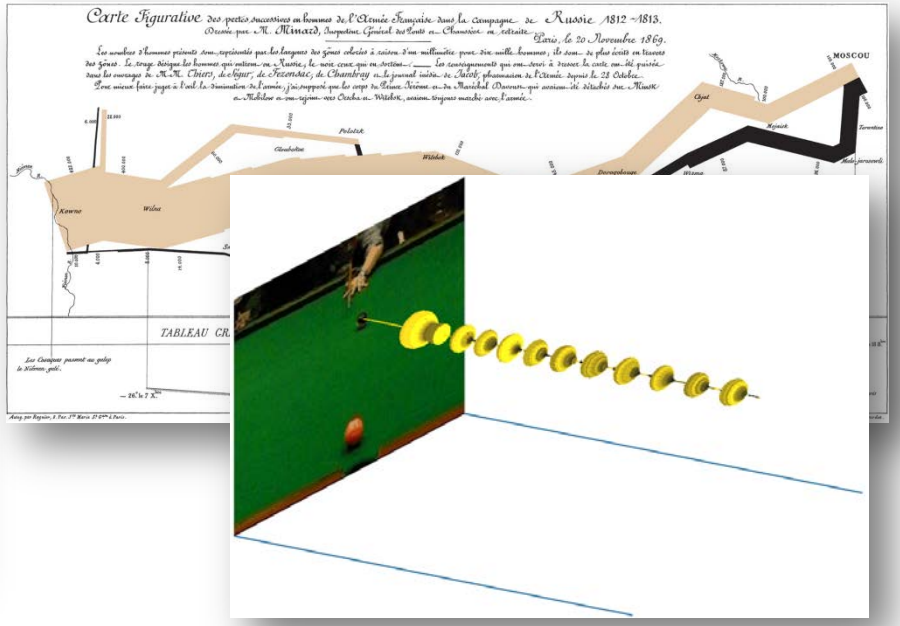


[TLX]

Snooker Skill Training[Höferlin10]

- Video visualization to improve snooker skills
- Validation meeting with 5 potential users
  - Introduction of visualization results with other standard visualizations on slides

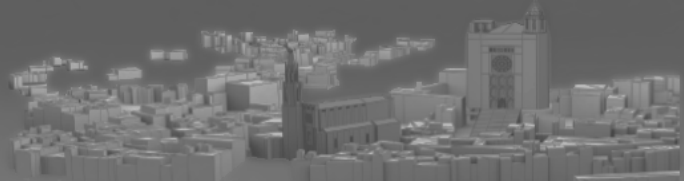




Snooker Skill Training[Höferlin10]

- 6 sets of questions and open discussion
  - Set 1
    - How do you perfect a particular cuing action by yourself?
  - Set 2
    - How do you teach an intermediate level player to perfect a particular cuing action?
    - How do you normally recognize a cuing problem?
    - How do you normally identify and record the progress of a player?
  - Set 3
    - Can you find any difference in the videos?
    - How long can a coach afford to spend time with an intermediate player in watching videos and discussing videos in order to identify problems and solutions?
    - How many video-based analysis would a coach be willing to go through each day?
    - How long would a player be willing to go through such a process with the coach?
  - Set 4
    - Are you familiar with this 2D visualization? (Minard's Map)
  - Set 5
    - Can coaches and players learn to recognize problems and progress from such visualizations?
    - Would you be happy to use video visualization to replace watching videos (a) completely, (b) mostly and to avoid watching videos repeatedly, (c) occasionally, (d) not at all, (e) other (to be specified)?
  - Set 6
    - Any other comments and suggestions?

What to Consider?[Frary03]



- Only relevant questions
  - Avoid annoyance and frustration
- Avoid leading questions
  - Don‘t imply certain responses
  - Example: „Don‘t you think...“
- Open-ended questions
  - Reveal unsuspected information
  - Willingness and ability to answer will vary
- Scales
  - Biploar scales, consistent labels
  - Uneven: Choice of neutral midpoint may have various reasons

Case Studies and Expert Evaluation

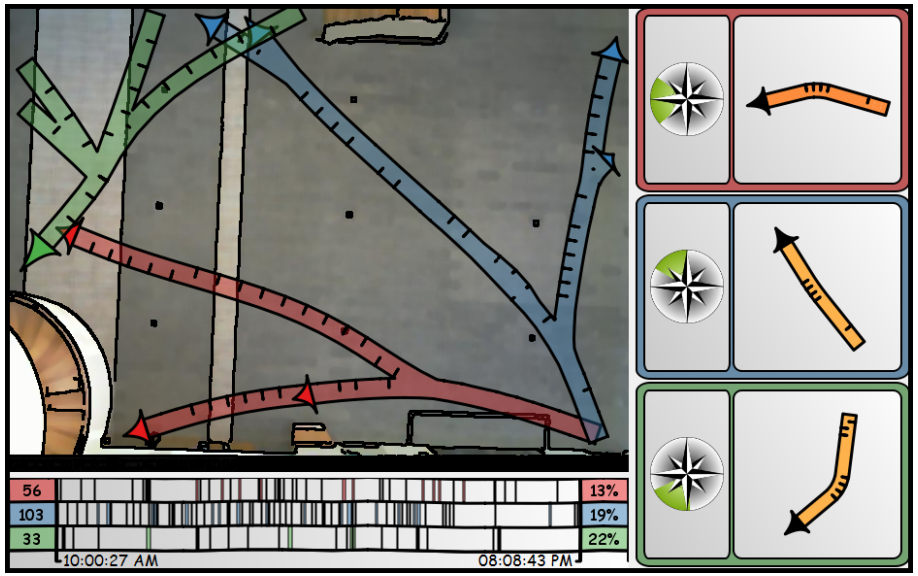
Think Aloud

- Expert review
  - Experts inspect the visualization
- Domain experts
  - Realistic application scenarios
- Information
  - Qualitative
  - Subjective
- Participants perform a task
  - Verbalizing thoughts, feelings, impressions
  - Verbalizations are protocolled

Example: Interactive Schematic Summaries

- Interactive Schematic Summaries for Faceted Exploration of Surveillance Video[Höferlin13]

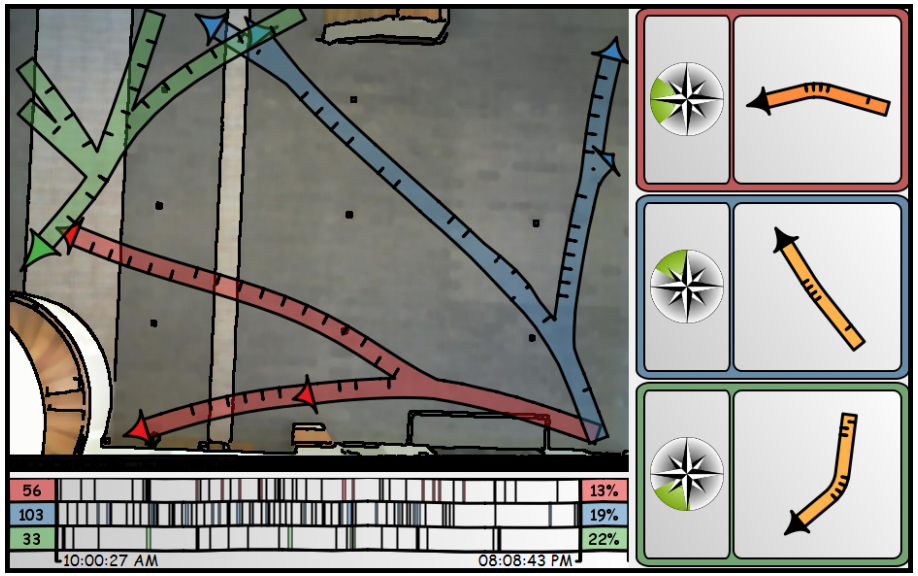
- Video exploration by trajectory browsing
- Initial user feedback with 5 experts
- Introduction with example
- **Task:** Analyzing typical movements in the data
  - Example: When are many people leaving the building?



[Höferlin13]

Example: Interactive Schematic Summaries

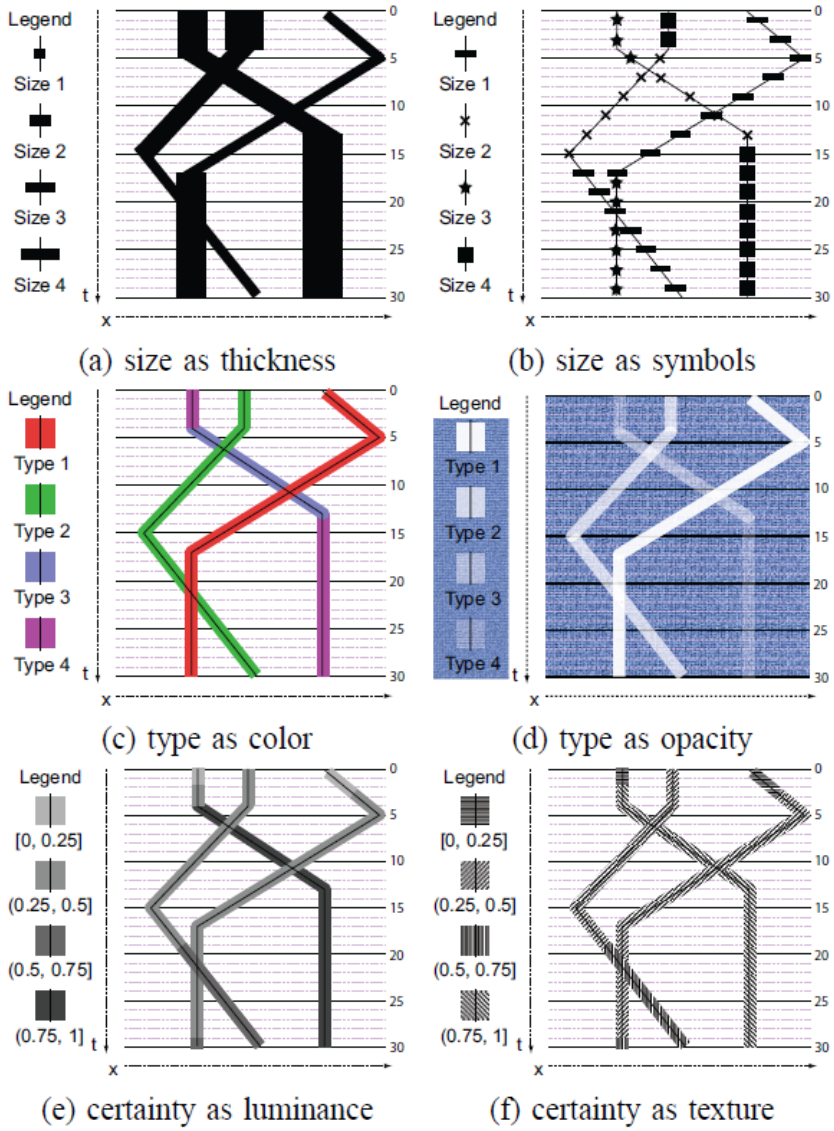
- Experts use the tool
  - Think aloud with audio protocol
- Participants verbalized:
  - What am I doing?
  - Why am I doing it?
  - What is noteworthy?
- Semantically matching comments between participants:
  - Useful to find prominent directions
  - More initial training needed
  - History of browsing steps



[Höferlin13]

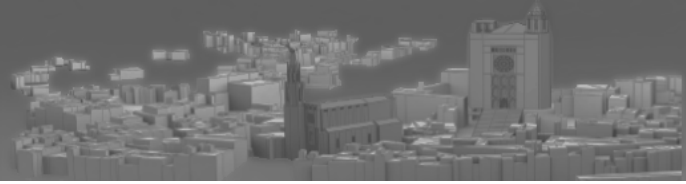
Example:
Action-Based Multifield Video Visualization [Botchen08]

- Survey on visual mappings
  - 3 attributes
  - 6 different mappings
- Rated by
  18 visualization experts
- Color and thickness are most favored mappings



What to Consider? [Lewis93] [VanSomeren94]

- Think aloud can disturb the cognitive process
  - Realism of the experiment biased
- Interpreation of results is subjective
  - Unambiguous answers
- Motivation can vary during the task
  - Remind participants to comment (neutral prompts)
  - Constructive interaction (pair testing)
- Task of appropriate difficulty
  - Don‘t choose an automatically solvable task



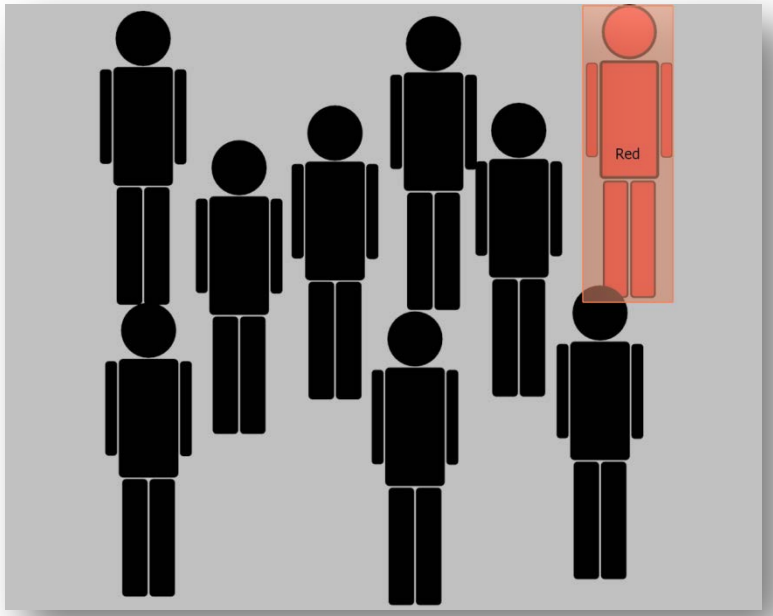
Eye-Tracking with Video Stimuli

What Can You Measure?

- Controlled lab study

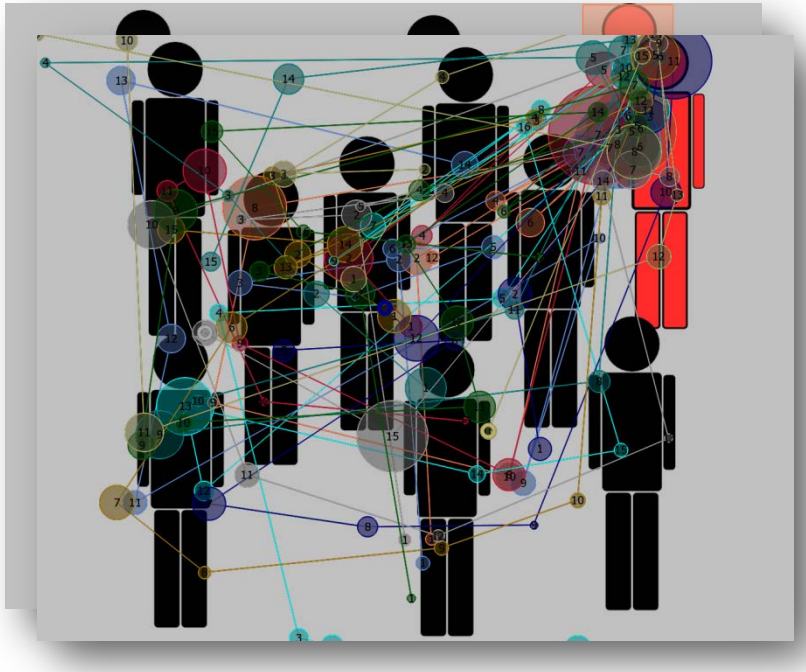
- Raw data with many information (left/right eye)
  - Gaze points (e.g., in monitor coordinates)
  - Distance
  - Pupil size
  - Fixations (with filtering)

- Different analysis methods
  - AOI – Areas Of Interest
  - Scan paths
  - Heat maps
  - Focus maps



What Can You Measure?

- Controlled lab study
- Raw data with many information (left/right eye)
  - Gaze points (e.g., in monitor coordinates)
  - Distance
  - Pupil size
  - Fixations (with filtering)
- Different analysis methods
  - AOI – Areas Of Interest
  - Scan paths
  - Heat maps
  - Focus maps

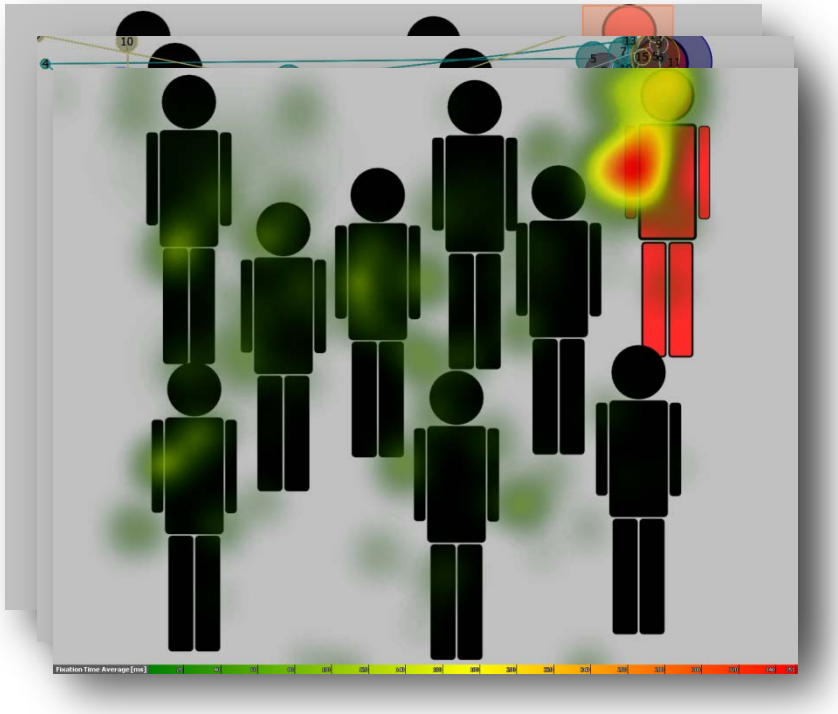


What Can You Measure?

- Controlled lab study

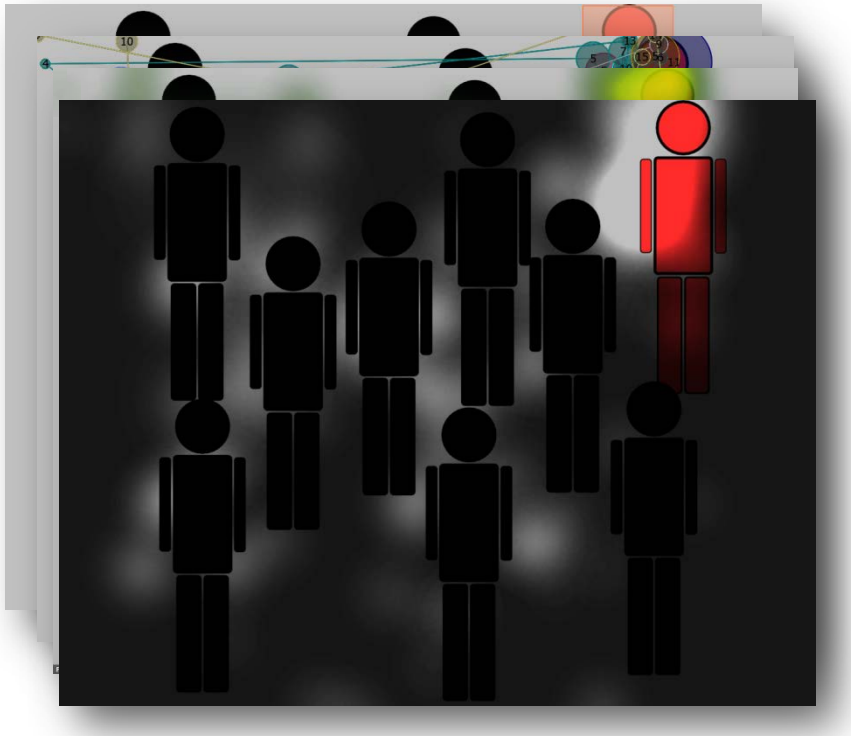
- Raw data with many information (left/right eye)
  - Gaze points (e.g., in monitor coordinates)
  - Distance
  - Pupil size
  - Fixations (with filtering)

- Different analysis methods
  - AOI – Areas Of Interest
  - Scan paths
  - Heat maps
  - Focus maps



What Can You Measure?

- Controlled lab study
- Raw data with many information (left/right eye)
  - Gaze points (e.g., in monitor coordinates)
  - Distance
  - Pupil size
  - Fixations (with filtering)
- Different analysis methods
  - AOI – Areas Of Interest
  - Scan paths
  - Heat maps
  - Focus maps

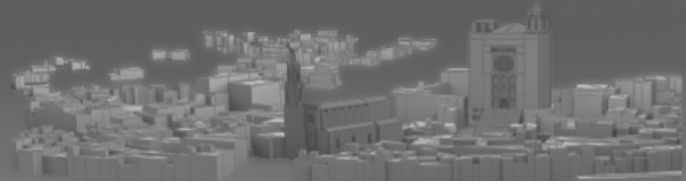


Eye-Tracking with Video Stimuli

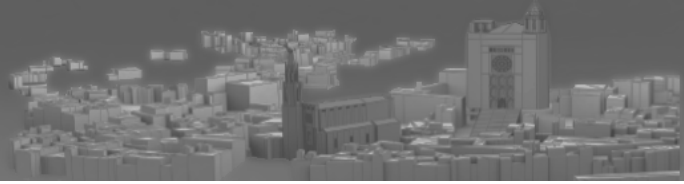
- Eye-tracking of videos
  - Often just visualization of gaze replay
  - Heat maps and scan paths not as informative as with static images
  - Dynamic AOIs needed

[ILIDS]

Problems



- Measurement issues
  - Glasses, contacts
  - Head position
- Dynamic areas of interest
  - Inaccuracy possible
  - Fixation coordinates + foveal region
- Fast moving objects
  - Latency possible



Hints for a Successful Study

Biases

Handle biases as well as possible

- Remove distractions
  - Cell phones off
  - Only relevant equipment on the table

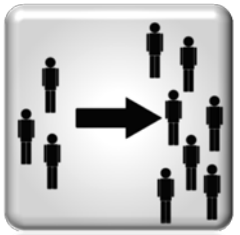
- Don‘t influence your participants
  - Careful answers to questions
  - No leading questions

- Don‘t underestimate exhaustion of the participants
  - Plan resting periods

Pilot Studies

Perform pilot studies

- Smaller number of participants



- Gain information to refine your design:
  - Are the results reasonable?
  - Is the task too easy/hard for reasonable results?
  - How long does it take for one participant to finish?

- If possible, refine and pilot again

Conducting a Study

- Write a schedule
  - Stepwise instructions
  - Replicability

- Introduce your visualization
  - Tutorial with example
  - Training

- Background information
  - Classification of participants
  - Participants remain anonymous

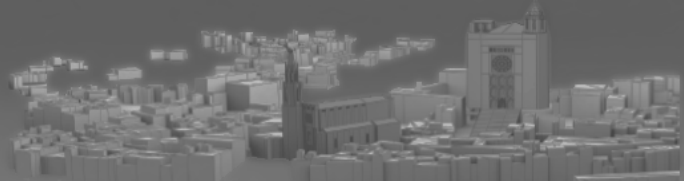
Ethics[Dumas08]

- Informed Consent
  - Provide information about study procedure
  - Option to quit at any point
  - Let participants read and sign

- Compensation?

- Keep information confidential
  - Random IDs for participants
  - Restricted use of data

Other Tutorials



Tutorials, Workshops and Courses on Evaluation

- VisWeek 2012 Workshop
  BELIV 2012 – Beyond Time and Errors: Novel Evaluation Methods for Visualization
  http://www.beliv.org/
- Eurographics 2011 Tutorial
  Scientific Evaluation in Visualization
- SIGGRAPH 2009 Course
  The Whys, How Tos, and Pitfalls of User Studies
  doi:10.1145/1667239.1667264

References

- [Botchen08] Botchen, R.; Schick, F.; Ertl, T.: Action-Based Multifield Video Visualization. *Visualization and Computer Graphics, IEEE Transactions on, 14*, pp. 885 -899, **2008**
- [Carpendale08] Carpendale, S.; Kerren, A.; Stasko, J.; Fekete, J.D.; North, C.: Evaluating Information Visualizations. *Information Visualization, Springer Berlin Heidelberg, 4950*, pp. 19-45, **2008**
- [Chen06] Chen, M.; Hashim, R.; Botchen, R.; Weiskopf, D.; Ertl, T.; Thornton, I.: Visual Signatures in Video Visualization *Visualization and Computer Graphics, IEEE Transactions on, 12*, pp. 1093 -1100, **2006**
- [Dumas08] Dumas, J.; Loring, B.: Moderating usability tests: principles and practice for interacting. *Morgan Kaufmann,* **2008**
- [Field03] Field, A.; Hole, G.: How to design and report experiments. *Sage publications London,* **2003**
- [Forsell10] Forsell, C.: A Guide to Scientific Evaluation in Information Visualization *Information Visualisation (IV), 14th International Conference*, pp. 162 -169, **2010**
- [Forsell12] Forsell C.; Cooper M.: A guide to reporting scientific evaluation in visualization. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 608-611, **2012**
- [Frary03] Frary, R.: A brief guide to questionnaire development *Virginia Polytechnic Institute & State University. Retrieved October, 7*, **2003**
- [Hart88] Hart, S.; Staveland, L.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research *Human mental workload, 1*, pp. 139-183, **1988**
- [Höferlin10] Höferlin, M.; Grundy, E.; Borgo, R.; Weiskopf, D.; Chen, M.; Griffiths, I. W.; Griffiths, W.: Video Visualization for Snooker Skill Training. *Computer Graphics Forum 29, 3, pp. 1053-1062,* **2010**
- [Höferlin11] Höferlin, B.; Höferlin, M.; Weiskopf, D.; Heidemann, G.: Information-Based Adaptive Fast-Forward for Visual Surveillance. *Multimedia Tools and Applications 55 (1)*, pp. 127-150 , **2011**
- [Höferlin12] Höferlin, M.; Kurzhals, K.; Höferlin, B.; Heidemann, G.; Weiskopf, D.: Evaluation of Fast-Forward Video Visualization *IEEE Transactions on Visualization and Computer Graphics (TVCG), 18*, pp. 2095-2103, **2012**
- [Höferlin13] Höferlin, M.; Höferlin, B.; Heidemann, G.; Weiskopf, D.: Interactive Schematic Summaries for Faceted Exploration of Surveillance Video, *IEEE Transactions on Multimedia,* **2013**, (to appear)
- [ILIDS] http://www.homeoffice.gov.uk/science-research/hosdb/i-lids/
- [Lewis93] Lewis, C.; Rieman, J.: Task-centered user interface design: A practical introduction, **1993**
- [TLX] http://humansystems.arc.nasa.gov/groups/TLX/downloads/TLXScale.pdf
- [VanSomeren94] Van Someren, M.; Barnard, Y.; Sandberg, J. and others: The think aloud method: A practical guide to modelling cognitive processes *Academic Press London,* **1994**