Multimedia analysis of video collections: visual exploration of presentation techniques in ted talks

A. WU AND H. QU. MULTIMODAL ANALYSIS OF VIDEO COLLECTIONS: VISUAL EXPLORATION OF PRESENTATION TECHNIQUES IN TED TALKS. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, 2018.

MARJANE NAMAVAR
UNIVERSITY OF BRITISH COLUMBIA

INFORMATION VISUALIZATION

FALL 2019

Motivation

What are some features (verbal/non-verbal) of a good presentation?

- Avoid incessant hand movements
- Don't leave hands idle

Problems

- Suggestions are puzzling learners
- Non-verbal presentation techniques has been neglected in large-scale automatic analysis
- Lack of research on the interplay between verbal and non-verbal presentation techniques
- Only limited data-mining techniques for existing research

Proposed Solution

- Quantitative analysis on the actual usage of presentation techniques
- In a collection of good presentations (TED Talks)
- To gain empirical insight into effective presentation delivery

Contributions

- A novel visualization system to analyze multimodal content
- Temporal distribution of presentation techniques and their interplay
- A novel glyph design
- Case study to report the gained insights
- User study to validate usefulness of the visualization system

Challenge

Multimodal content

- Frame images
- Text
- Metadata

User-Centered Design Process

[Fig. 2. A. Wu and H. Qu. Multimodal analysis of video collections: Visual exploration of presentation techniques in ted talks. IEEE Transactions on Visualization and Computer Graphics, 2018.]

Preliminary Stage

Contextualized Interview

- Three domain experts
- Individual interviews to understand main problems
- Problems:
  - Case-based evidence rather than large-scale automatic analysis
  - Manual search to find examples

Preliminary Stage

Focus Group

- Before:
  - 14 Candidates
  - Mentioned in the domain literature
  - Quantifiable by computer algorithms
- After:
  - Three very significant and feasible presentation techniques
  - Rhetorical modes
  - Body postures
  - Gestures

Preliminary Stage

Presentation techniques

1) Rhetorical mode
- Narration
- Exposition
- Argumentation

2) Body Posture
- Close Posture
- Open Arm
- Open Posture

3) Body Gesture
- Stiff
- Expressive
- Jazz

Iteration Stage

- Three rounds
- Paper-based design and code-based prototyping
- Feedback-based enhancement

Analytical Goals

G1: To reveal the temporal distribution of each presentation technique

G2: To inspect the concurrences of verbal and non-verbal presentation techniques

G3: To identify presentation styles reflected by technique usage and compare the patterns

G4: To support guided navigation and rapid playback of video content

G5: To facilitate searching in video collections

G6: To examine presentation techniques from different perspectives and provide faceted search

Visualization Tasks

T1: To present temporal proportion and distribution of data

T2: To find temporal concurrences among multimodal data

T3: To support cluster analysis and inter-cluster comparison

T4: To compare videos at intra-cluster level

T5: To enable rapid video browsing guided by multiple cues

T6: To allow faceted search to identify examples and similar videos in video collections

T7: To display data at different levels of detail and support user interactions

T8: To support selecting interesting data or feature space

T9: To algorithmically extract meaningful patterns and suppress irrelevant details

System Architecture

- **Data Processing**

  Collect TED talks and extract presentation techniques

- **Visualization**

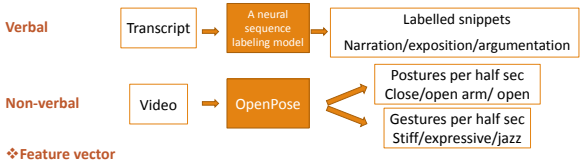
  Interactive visual analytic environment for deriving insights

[Fig. 3. A. Wu and H. Qu. Multimodal analysis of video collections: Visual exploration of presentation techniques in ted talks. IEEE Transactions on Visualization and Computer Graphics, 2018.]

Data Processing

- **Data**
  - 146 TED talks gathered from the official website in the chronological order
  - Videos
  - Transcript (segmented into snippets with various time intervals)
  - Metadata
- **Data processing techniques**
  - Verbal
  - Non-verbal

Data Processing (cont.)



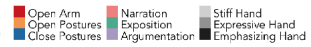
❖**Feature vector**
- 9x1 vector
- Temporal proportion of each of the nine techniques

Visual Design

[Fig. 5. A. Wu and H. Qu. Multimodal analysis of video collections: Visual exploration of presentation techniques in ted talks. IEEE Transactions on Visualization and Computer Graphics, 2018.]

Unified Color Theme

- Posture: Cool color for close posture
- Gesture: higher saturation for larger movement
- Rhetorical mode: Color psychology
  - Narration: Pink (Symbolizing life)
  - Exposition: Green (Reliability)
  - Argumentation: Purple (Wisdom)



[Part of Fig. 7. A. Wu and H. Qu. Multimodal analysis of video collections: Visual exploration of presentation techniques in ted talks. IEEE Transactions on Visualization and Computer Graphics, 2018.]

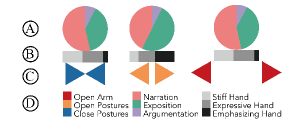
TED talk glyph

Metaphor of the human body

Head: Pie-chart, proportion of rhetorical modes

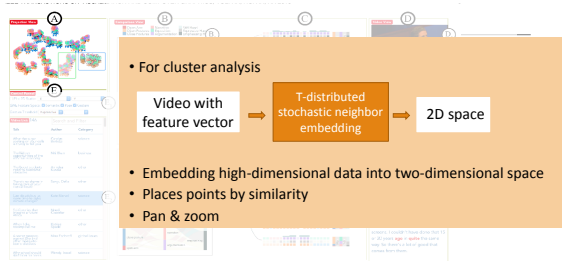
Shoulders: Bar-chart, percentage of gestures

Triangles: Frequent hand posture



[Fig. 7. A. Wu and H. Qu. Multimodal analysis of video collections: Visual exploration of presentation techniques in ted talks. IEEE Transactions on Visualization and Computer Graphics, 2018.]

Projection View

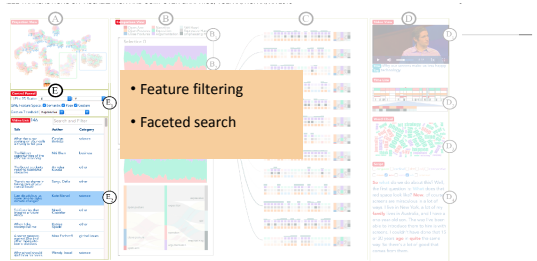


- For cluster analysis

Video with feature vector → T-distributed stochastic neighbor embedding → 2D space

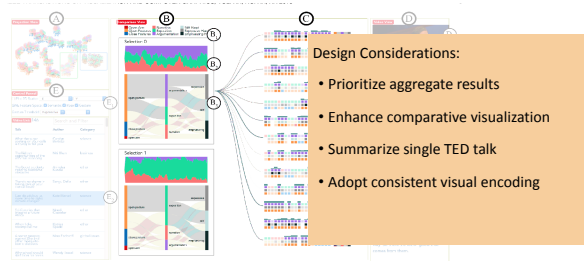
- Embedding high-dimensional data into two-dimensional space
- Places points by similarity
- Pan & zoom

Control Panel



- Feature filtering
- Faceted search

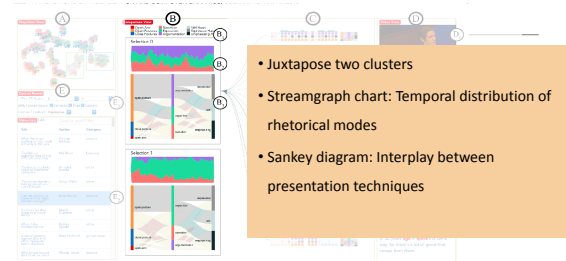
Comparison View



Design Considerations:

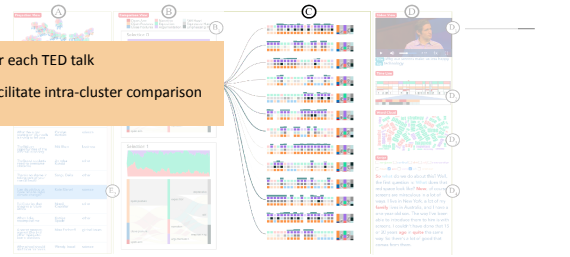
- Prioritize aggregate results
- Enhance comparative visualization
- Summarize single TED talk
- Adopt consistent visual encoding

Comparison View -> Aggregate View



- Juxtapose two clusters
- Streamgraph chart: Temporal distribution of rhetorical modes
- Sankey diagram: Interplay between presentation techniques

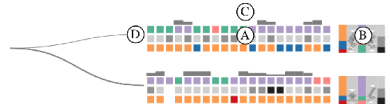
Comparison View -> Presentation Fingerprinting



- For each TED talk
- Facilitate intra-cluster comparison

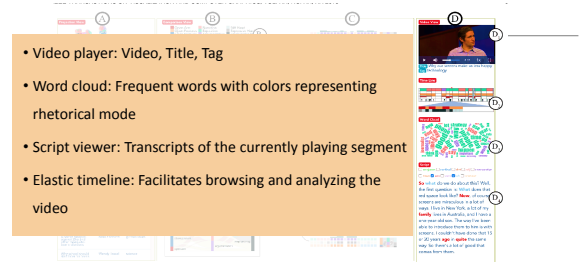
Comparison View -> Presentation Fingerprinting(cont.)

- Rows (top to bottom): Rhetorical mode, Gesture, Posture
- Uniform time interval of 5% of the talk duration
- Embedded bar-chart: Top concurrence tuples



[Fig. 9. A. Wu and H. Qu. Multimodal analysis of video collections: Visual exploration of presentation techniques in ted talks. IEEE Transactions on Visualization and Computer Graphics, 2018.]

Comparison View -> Video View



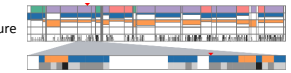
- Video player: Video, Title, Tag
- Word cloud: Frequent words with colors representing rhetorical mode
- Script viewer: Transcripts of the currently playing segment
- Elastic timeline: Facilitates browsing and analyzing the video

Elastic Timeline

- Two layers
- First layer: Timeline is segmented according to the transcript snippet
- Usage of presentation techniques arranged vertically
- Row 1: Rhetorical mode
- Row 2-4: Three types of body posture
- Bar-charts: The proportion of corresponding posture during the time interval
- Row 5: Bar-chart represents body gesture

Unfold the bottom layer

- Gestures and postures during the selected segment
- Each grid show a half second
- Blank grid: Any information is non-retrievable



[Fig. 10. A. Wu and H. Qu. Multimodal analysis of video collections: Visual exploration of presentation techniques in ted talks. IEEE Transactions on Visualization and Computer Graphics, 2018.]

Evaluation -> Case Study

- With 3 experts and 3 students
- To reflect the fulfillment of analytical goals and gain insight
- Used the system and provided feedback

- Results:
  - System reached the analytical goals
  - Findings matched the theories
  - Incorporate the system into theirs current research and teaching practices
  - Suggested more gestures such as pointing

Evaluation -> User Study

- With 16 students
- To demonstrate the capacity of undertaking visualization tasks and gather feedback
- Went through a series of tasks and provided feedback

- Results:
  - All participants understood and completed tasks
  - They agreed system is usable for video collections
  - Less satisfied with video comparison view

Limitations and Future Work

LIMITATIONS

- Research Scope
- Accuracy
- Presentation Fingerprinting
- Overlapping among glyphs
- Comparison of two clusters

FUTURE WORK

- Extract additional features
- Improve accuracy
- Assist more analytical tasks
- Evaluate with other presentation scenarios

Analysis Summary

- What (data):
  - Video (image frames)
  - Text (transcripts)
  - Metadata (tags)
- What (derived):
  - Tags for postures per half sec/gestures per half sec/rhetorical mode per snippet
  - Feature vector (temporal proportion of nine techniques)
- Why (tasks):
  - T1-T9

Analysis Summary (cont.)

- How (encode):
  - 2D plot
  - Bar-chart
  - TED talk glyph (using pie-chart, bar-chart, distance and direction of triangles)
  - Streamgraph
  - Sankey diagram
  - Links (relation between each talk and aggregated data)
  - Table (each talk)
  - Grid (timeline)
  - Stacked bar-chart (postures in timeline)
  - Consistent color-map(hue/saturation)
- How (Reduce):
  - Filtering of features
  - Aggregation

Analysis Summary (cont.)

- How (Facet):
  - Partition into multiform views
  - Juxtapose views for comparison
  - Linked highlighting
  - Linked navigation
  - overview–detail with selection in overview populating detail view
- How (Manipulate):
  - Select (clusters, control panel & video view)
  - Collapse and expand
  - Zoom & pan (projection view)

Critique

STRENGTHS

- Carefully designed with well justified design choices
- Sophisticated view coordination (screen-space effective & different levels of details)
- Consistency in visual mappings
- Reduce cognitive/memory burden
- Carefully designed glyph
- Inter-, Intra-cluster & within-video analysis

WEAKNESSES

- Why TED talks / Which TED talks
- Evaluated only on a small set of TED talks
- Some parts are not related to any of the tasks (word cloud)
- Does not discuss the ability of the system to scale when number of features or videos or the duration of videos increases
- Only captures simple relationships among presentation techniques
- Unnecessary encodings / details without explanation (elastic timeline)