Videotater: An Approach for Pen-Based Digital Video Segmentation and Tagging

Evan Turner
CAP 5937 – Pen-Based Interfaces

Videotater – Editing Interface
- Display video
  - “Polyfocal”
- Segmentation
- Tagging
- Improvements for novices and experts
  - “Video repurposing”
Traditional Video Editing

- Adobe Premiere

Source: [2]

Advances

- Precise location and execution of segments
  - “Magnetic lasso”
- Refining/trimming rough segments
- Different modalities for pen-based interaction
  - Space
  - Time
  - Pressure
Tangible Interaction Modality

- Low pressure: indicates selection
- High pressure: indicates tagging
- Pen tip applies tags, pen eraser removes them
Spatial Interaction Modality

- Gestures on timeline:
  - Vertical line: split segment
    - Mean x-value of line or nearest likely boundary
  - Line between segments: merge

Tagging

- On segment: applies tag to entire segment
- On stripe image: applies tag only to frames touched
- Allows for shot- or frame-level tagging
- Can be removed by dragging eraser over timeline
Polyfocal Visualization

Interaction with Polyfocal Visualization

- Modeless during mouse or pen hovering
- Visualization tracks the cursor
- Clicking or dragging engages a mode
  - Dismissed by clicking again
- Context images:
  - Hover: Displays at larger size
  - Click: New in- or out-point for segment
Implementation

- Java Swing
- Video is pre-processed
  - ~20% of real-time on 2.1 GHz Pentium M
  - Generate/compress thumbnails

User Feedback Setup

- Three experienced video editors
- Segment and tag two minute video while thinking aloud
- Given demonstration
- Allowed to explore interface before think-aloud
User Feedback Pros

- Timeline interface “compelling”
- Hovering/stripe enhanced ease of segmenting and tagging
- “Informal timings of segmentation efficiency indicates it takes roughly 2x real-time while maintaining high precision.”
- Pressure-modality intuitive, quick

User Feedback Cons

- Split gesture “felt sloppy”
- Subsequent trimming redundant
- Switching to eraser “natural”, but slow
- Polyfocal interface difficult to parse
- Vertical frame listing confusing
References


Questions?