FXPAL Interactive Search for TRECVID 2004

John Adcock, Matthew Cooper, Andreas Girgensohn, Lynn Wilcox
Overview

• First time doing search
  – 2\textsuperscript{nd} year of participation overall

• Emphasis on interface elements
  – Rich visualization of search results
  – Quick and easy exploration of results

• Straightforward search engine
  – Text search over ASR transcripts
    • Literal search with Lucene
    • Fuzzy search with LSS
  – Keyframe search by image similarity
    • Color correlograms
Preprocessing

Unit of search retrieval is a “story”, but we couldn’t don’t have reference story segmentation for the test set

- Group reference shots into “stories”
  - Bootstrap an LSS with common shot boundaries and ASR
  - use similarity-matrix method to find “story” boundaries
- Given new story boundaries
  - Generate text indices for story and shots
  - Generate story-based LSS for search
Search Engine

• User specifies combination of:
  – Text query
    • Literal query using Lucene or fuzzy query using LSS
  – Image examples
    • Any keyframe in the interface can be dragged onto the image example area
  – Text/image weighting is static and equal
  – Max image similarity of shot propagated to story
  – Text similarity of story propagated to shot
    • Averaged with shot-based text similarity
Search Engine

- Query text
- Lucene Search
- LSS Search
- Image Color Correlogram Search
- Combine
- Searcher option
- Ranked Stories
Interface Elements

- Stories summarized in keyframe “quads”
- Navigate through stories to video timeline/shots
- Transparent icon overlays
  - Visited: grayed
  - Relevant: green
  - Irrelevant: red
- Query-relevance shown with size and color
- Hotkeys for most actions
- Multi-select and drag and drop
Story Summary Quads

- Query-dependent story summary
  - Use 4 highest scoring shots in the story
  - Allocate space proportional to score

Story thumbnail

Shot thumbnails
Building on searches

- Find similar
  - Use shot/story text for search

- Add related
  - Auto re-query with existing results

12/06/1998 CNNa 0:22 0.6
insight circuit henry hyde
lawyers committee judiciary impeachment

Add to result
No result
Undo no result
Find similar

Add related
Expanded Story / Timeline Browsing

- Selecting a story expands the video at that point
  - Clickable video timeline with relevancy shading
  - Clickable story quad timeline
  - Shot thumbs marked with relevancy
  - Overlay on shots marked (non)relevant
  - Mouse-overs zoom in the media player and tool-tip shows relevancy context
  - Double clicks play video in the media player
Experiments

• 6 searchers answering 12 topics each in latin square
  – Pairs of orthogonal users grouped together
    • Each topic answered 3 times
  – Searchers include 2 primary developers
    • 1 ended up in best and 1 in worst performing group

• Each of the 3 complete searcher runs goes through 3 “systems” or methods for filling out the shot list yielding 9 total submissions
System Types

- **Type 1:**
  re-issue user queries and weight results of each query by precision against the user-labeled shots

- **Type 2:**
  take text from all relevant shots and issue a single new LSS-based text query

- **Type 3:**
  take text from each relevant shot in turn for LSS-based query and apply query ranking as in system type 1

*Shots marked as not-relevant excluded from system results*

*Every system type preceded by bracketing the user-retrieved shots*
Submissions

User IDed Shots

Bracketed Shots

System1 (Weighted)

System2 (LSA1)

System3 (LSA2)
Results

- Ranks 3-6, 9-13 in overall MAP
  - Strongly user dependent (user groups clump together)
  - Post-processing methods perform nearly same
User vs. System

System Summary

MAP

Group 1 | Group 2 | Group 3

User Group

WEIGHTED | LSA1 | LSA2 | Bracketed | None

FX Palo alto Laboratory Inc @ trecvid 2004
User vs. System in Overall

MAP

Complete submission
With bracketing
User selected only
Other contributors

Submission
Performance by Question

- Overall median
- FXPal average
- Overall max

Items: people on steps or stairs, pedestrians and vehicles, bicycles rolling, umbrellas, fingers striking keyboard, buildings on fire, handheld weapon firing, golf ball into the hole, Bill Clinton, horses in motion, people and dogs, wheelchairs, signs at a protest, zooming in US Capitol dome, Benjamin Netanyahu, buildings with flood waters, Henry Hyde, Saddam Hussein, Sam Donaldson, hockey rink, Boris Yeltsin.
Directions

• More sophisticated:
  – Story segmentation
  – Image similarity / video features
• Simplify user interface for non power-users and more typical search and re-use tasks
• Handle multiple simultaneous media streams
  – Presentation slides
  – Multi-camera capture