Learning From Video Browse Behavior

TRECVID 2009
Problem Statement

Starting results relatively weak
- Combination of query methods troublesome

Possible solutions:
- Optimize result selection
- Visualize multiple query methods simultaneously
- Analyze user browse behavior
Optimize Result selection?
Focus + Context browsing

focus shot
context

START → inspect focus → choose label → inspect navigation → choose option → STOP
Focus + Context browsing

Focus:
- defined by the current focal shot

Context:
- defined by the rest of the interface
- We use: multi thread browsing
  
  A thread is a linked sequence of shots in a specified order, based upon an aspect of their content
Threads used

query threads
- merged result of query-by-text and/or query-by-concept and/or query-by-example

time threads
- based on the shots in the video containing the focal shot

visual threads
- based on visual similarity of focal shot

history thread
- based on the previous user browse behavior
Multi Thread Browsing: ForkBrowser

- visual thread
- query thread
- visual thread
- focal shot
- time thread
- history thread
Multi Thread Browsing: ForkBrowser
Learning From Video Browse Behavior

Problem Statement

Starting results relatively weak
- Combination of query methods troublesome

Possible solutions:
- Optimize result selection
  - We propose: Focus + Context
- Visualize multiple query methods simultaneously
  - We propose: Multi Thread Browsing
- Analyze user browse behavior
  - We propose: Relevance Feedback based on context
Relevance Feedback based on Context

Based on online SVM learning

- User provides positive annotations
- System gathers negative annotations based on user browse behavior
  - using displayed context

User switches query thread when current results seem exhausted
Relevance Feedback based on Context

All displayed shots accumulate a score to have been seen by the user.

When a shot reaches a threshold that shot is used as a negative example.
How to evaluate performance?

Problem with measuring real world users

Component level evaluation requires user simulation
User Simulation with a State Machine

START
inspect center shot
playback center shot

label center shot

inspect navigation threads

enough results found
STOP

most relevant shots visible in thread $\tau_k$
move into $\tau_k$

no relevant shots visible in any thread
move into $\tau_0$

relevant shots visible in previous state
move into $\tau_a$

no relevant shots visible in any thread after $N$ interaction steps
use relevance feedback to generate new $\tau$
Learning From Video Browse Behavior

Experimental Setup

TRECVID 2008 dataset
- 200 hours of video
- 48 topics, with (incomplete) annotations
- 57 semantic concepts (21 of '08, 37 of '07)
- best concepts taken as optimal starting query

Experiment A:
What is the benefit of having multiple threads?

Experiment B:
When should a user switch to relevance feedback results?
Learning From Video Browse Behavior

**Experiment A**

**What is the benefit of having multiple threads?**

**Measure**

- retrieval performance vs number of shown threads
- number of positives after 500 actions, repeat for:
  - query only
  - query + time (CrossBrowser)
  - query + time + visual similarity (ForkBrowser)
Experiment A

Combined result plot to determine benefit of multiple threads

Topics:
- people with mostly trees and plants (0236)
- people (playing) with children (0239)
- piece(s) of paper with writing (0228)
- people sitting at a table (0222)
- a map (0231)
- vehicle(s) passing the camera (0230)
- food and/or drinks on a table (0241)
- people with body of water visible (0229)
- face filling more than half of frame (0227)
- a vehicle approaching the camera (0244)
- interviewed woman talking to cam (0237)
- black and white photographs (0233)
- road taken from a moving vehicle (0224)
- vehicle moving away from the cam (0234)
- people with horses (0223)
- people with books (0240)
- a bridge (0225)
- a person opening a door (0221)
- people walking into a building (0232)
- people sitting down in a chair (0242)
- waves breaking onto rocks (0236)
- person on the street talking to cam (0235)
- person pushing a child in a stroller (0238)
- people looking into a microscope (0243)
Experiment B

When should a user switch to relevance feedback results?

Measured

- optimal # of actions without results before using relevance feedback
Learning From Video Browse Behavior

Experiment B

- RF after 10 irrelevant
- RF after 15 irrelevant
- RF after 25 irrelevant
- RF after 50 irrelevant

For topics with a low baseline, RF has the most benefit.

The earlier relevance feedback is used the better.
TRECVID 2009 results

TRECVID 2009 Interactive Search Task Results

- a road taken from a moving vehicle through the front window
- a crowd of people, outdoors, filling more than half of the frame area
- a view of one or more tall buildings (more than 4 stories) and the top story visible
- a person talking on a telephone
- a closeup of a hand, writing, drawing, coloring, or painting
- exactly two people sitting at a table
- one or more people, each walking up one or more steps
- one or more dogs, walking, running, or jumping
- a person talking behind a microphone
- a building entrance
- people shaking hands
- a microscope
- two more people, each singing and/or playing a musical instrument
- a person pointing
- a person playing a piano
- a street scene at night
- printed, typed, or handwritten text, filling more than half of the frame area
- something burning with flames visible
- one or more people, each at a table or desk with a computer visible
- an airplane or helicopter on the ground, seen from outside
- one or more people, each sitting in a chair, talking
- one or more ships or boats, in the water
- a train in motion, seen from outside
- camera zooming in on a person's face

Inferred Average Precision

concept detectors
visual threads
relevance feedback
Conclusions

Results indicate:

- showing multiple threads yield better performance
- also increases the time to perceive results for real world humans
- We found a inverse correlation between # of threads shown and importance of initial query
- Relevance Feedback yields greatest benefit for topics which would otherwise have limited results.

ForkBrowser Focus + Context browsing paradigm, together with good initial concepts, consistently performs well
Any questions?