CuZero: Interactive Video Search for Informed User

Shih-Fu Chang
(joint work with Eric Zavesky)

Digital Video and Multimedia Lab
Columbia University
www.ee.columbia.edu/dvmm
Oct. 2008
The Goggle Era Search Paradigm

“…type in a few words at most, then expect the engine to bring back the perfect results. More than 95 percent of us never use the advanced search features most engines include, …”


- Simple keyword search is the primary search method
- Extension to visual search seems to be a natural desirable step
- Examples: text search on Google, Flickr, Youtube
The Role of User

Formulate query

Result triage
An Example of Web Image Search

- Text Query: “Manhattan Cruise” over Goggle Image

How to visualize large image/video sets?
Why are these images returned? No word-document matching
If this does not work, how do we choose better search terms?
Minor Changes in Keywords → Big Difference

- Text Query: “Cruise around Manhattan”
Pains of Frustrated Users

- Forced to take “one shot” searches, iterating queries with a trial and error approach...
Problem:
User’s Inability in Forming Queries

- Difficult to choose words/concepts without in-depth knowledge of data and vocabulary

Example query: Find shots of something (e.g. vehicle, aircraft, building, etc) on fire
A Call for Research Attention

Research needed to address user frustration

A lot of work on content analytics
Solution--
Keep User Informed and Engaged

- Help users stay “informed”
  - in formulating queries
  - in understanding search results
  - in changing search rapidly and flexibly

- Keep users “engaged”
  - Instant feedback with minimal disruption as opposed to slow “trial-and-error”
An attempt: **CuZero**

Zero-Latency Informed Search & Navigation
A lot of efforts in visual tagging

- Multi-source tagging
  - Social tagging
  - Tagging by game
  - Geo-context tagging
  - Auto-tagging by content analysis

- Useful for keyword search
Tagging by Content Analytics

- Audio-visual features
- Machine learning models
- Web metadata
- Context (geo, user etc)

Large pools of semantic tags, concept labels

Semantic Indexes

- Anchor
- Snow
- Soccer
- Building
- Outdoor

Statistical models
Example: LSCOM concept pool/models

374 concept detection models: objects, people, location, scenes, events, etc

airplane airplane_takeoff airport_or_airfield armed_person building car
cityscape crowd desert dirt_gravel_road entertainment explosion_fire fire forest
highway hospital insurgents landscape maps military military_base
military_personnel mountain nighttime people-marching person
powerplants riot river road rpg shooting smoke tanks urban vegetation
vehicle waterscape_waterfront weapons weather
A User Community Approach to Vocabulary Development

Large Scale Concept Ontology for Multimedia
*(IBM-Columbia-CMU joint effort in 2005-6)*

- Ontology for tagging multi-source news video for analyst tasks
- Joint effort by government analysts, librarians, researchers
- Criteria: Useful, Observable, and Machine Detectable
- 834 concepts defined, extended to 2000+ by Cyc, 449 concepts annotated
- Labeled over TRECVID 2005 development set, 61,000 shots
  - 30+ annotators at Columbia and CMU
  - 33 million judgments
- Free for download (350+ downloads so far)
Results of sample concept detectors

- waterfront
- bridge
- crowd
- explosion fire
- US flag
- Military personnel

- Does not really solve the computer vision problem
- But the plurality of the models make it interesting for search
- Similar to speech retrieval from imperfect ASR
Sample detection results (TRECVID20)

- Airplane flying
- Classroom
- Demonstration or Protest
- Cityscape
- Singing
Example: “Smoke” Detector

- Compare to other tags, content-based analysis provide precision for deep tagging
- Ping point exact segment of concept
A prototype **CuZero**: Zero-Latency Informed Search & Navigation
Informed User:
Instant Informed Query Formulation
Informed User for Visual Search: Instant visual concept suggestion
Lexical mapping

Mapping keywords to concept definition, synonyms, sense context, etc
Co-occurrent concepts

Basketball courts and the American won the Saint Denis on the Phoenix Suns because of the 50 point for 19 in their role within the National Association of Basketball. George Rizq led Hertha for the local basketball game the wisdom and sports championship of the president. Baghdad to attend the game I see more goals and the players did not offer great that Beijing Games as the beginning of his brilliance Nayyouf 10 this atmosphere the culture of sports championship.
Query-Time Concept Mining

Initial text query

Find dominant visual concepts

visual mining

results

person

meeting

dominant concepts
CuZero Real-Time Query Interface (demo)

- Query auto completion & expansion
- Instant Concept Suggestion
Achieving Zero Latency Exploration

- Overlap (concept formulation) with (map rendering)
- Hide rendering latency during user interaction
- Course-to-fine concept map planning/rendering
- Speed optimization on-going …
A prototype **CuZero**: Zero-Latency Informed Search & Navigation

---

**ICQP** (Informed Concept Query Panel)

- **Query Topic**
- **Training Corpus**

**RtCNM** (Real-time Concept Navigation Map)

- **Future Work**
  - Alternate Concept Navigation Map

**Large Concept World** (e.g., LSCOM)

- car
- airplane
- urban
- fire
- building
- car crash
- road
- sky
- person

**Result Window**

---

Guided revision of query input text

Interactive concept selection

Instant inspection of query permutations

Real-time refresh
Address Information Overflow: Effective Visualization

linear browser restricts inspection flexibility

Informedia Concept Filter

only outdoor
Informed User:
Rapid Exploration of Results

Media Mill Rotor Browser
CuZero visualization:
Help users stay informed and engaged

- Create a concept map as search anchors
- Direct user control by sliding through the map
- Instant display for each location, without new query
Achieve Breadth-Depth Flexibility by instant concept map navigation (demo)

- Breadth: Quick sliding in the concept map
- Depth: Deep exploration of each result list
Columbia CuZero Video Search Engine
- addressing the search frustration

**Query Formulation**

Support **intuitive** formation in arbitrary **combination** of semantic concepts and visual examples.

**Navigation and exploration**

**Real-time** refinement and update of query.

**Fast, lightweight environment**

Web-based client-server approach for instant deployment with large data archive.

(Video Source: Youtube)
Simple Extensions

- More flexibility in modifying concept map
- Construct composite “super” concepts from simple ones
- Construct personal search toolbox
- Apply concept template to streaming mode
Concept Map Refinement
(current method requires multi-panel interaction)

1. User wants to replace ‘road’ with a related concept.
2. Co-occurrence can suggest related concepts.
3. Changing navigation map requires additional query reformulation and search execution.
4. Will interrupt the browsing task and requires cumbersome text-entry and manual concept selection.
In-place Anchor Editing

- Allow fast lookup of related anchors or allow user to swap in a ‘back-up’ anchor
- May also suggest strong metadata relationships
  - Capture time
  - Author/channel
  - Geo location
Formulating a Super-Anchor

- Navigation map allows precise specification of anchor weights
- For search topics of high-interest, user may have invested significant time in formulation/browsing
- With ideal location in navigation map, want to save this anchor configuration
  - Apply on new search at later time
  - Share with other users for collaborative filtering
  - Eventually build a personalized search tool box
Super-Anchors: Automated Alerts

- Passive monitoring of live streams to trigger alert
- Analyst benefits from customized search parameters for specific target
- Also reuse super-anchors in new navigation map.

![Diagram of Super Anchor and alternate database or streaming content threshold]

Super Anchor
(explosion + car + outdoor + road)

0.13264 Explosion Fire
0.00641 Outdoor
0.00641 Car
0.00031 Road

threshold
Deeper issues

- Is keyword-concept search the right paradigm?
- Perhaps for certain vertical domains?
- How far can automatic content analytics go?
  - Scale up the number of concepts?
  - Find the right visual lexicon?
  - Are they accurate enough for search?
Summary

- Effective visual search requires attention to help improve user experience
- More challenges than tagging and retrieval
- CuZero Design Principles
  - Keep user informed and engaged in
    - Query formulation and result visualization
  - Instant feedback
  - Support fusion of search modalities
    - Visual concept
    - Example
    - Keyword /speech transcripts
More Information

- LSCOM lexicon and annotation
  - Include 449 concepts (object, scene, event et al)
  - http://www.ee.columbia.edu/lscom

- Columbia DVMM Lab
  - http://www.ee.columbia.edu/dvmm
  - 374 SVM-based concept detectors
  - Video search demos