# TRECVID 2010 Known-item Search by NUS

Xiangyu Chen, Jin Yuan , Liqiang Nie, Zhengjun Zha, Shuicheng Yan
Tat-Seng Chua

National University of Singapore, Singapore

# Outline

- Introduction
- Auto Search
- Interactive Search
- UI of Our System & Demo
- Conclusion & Future Work

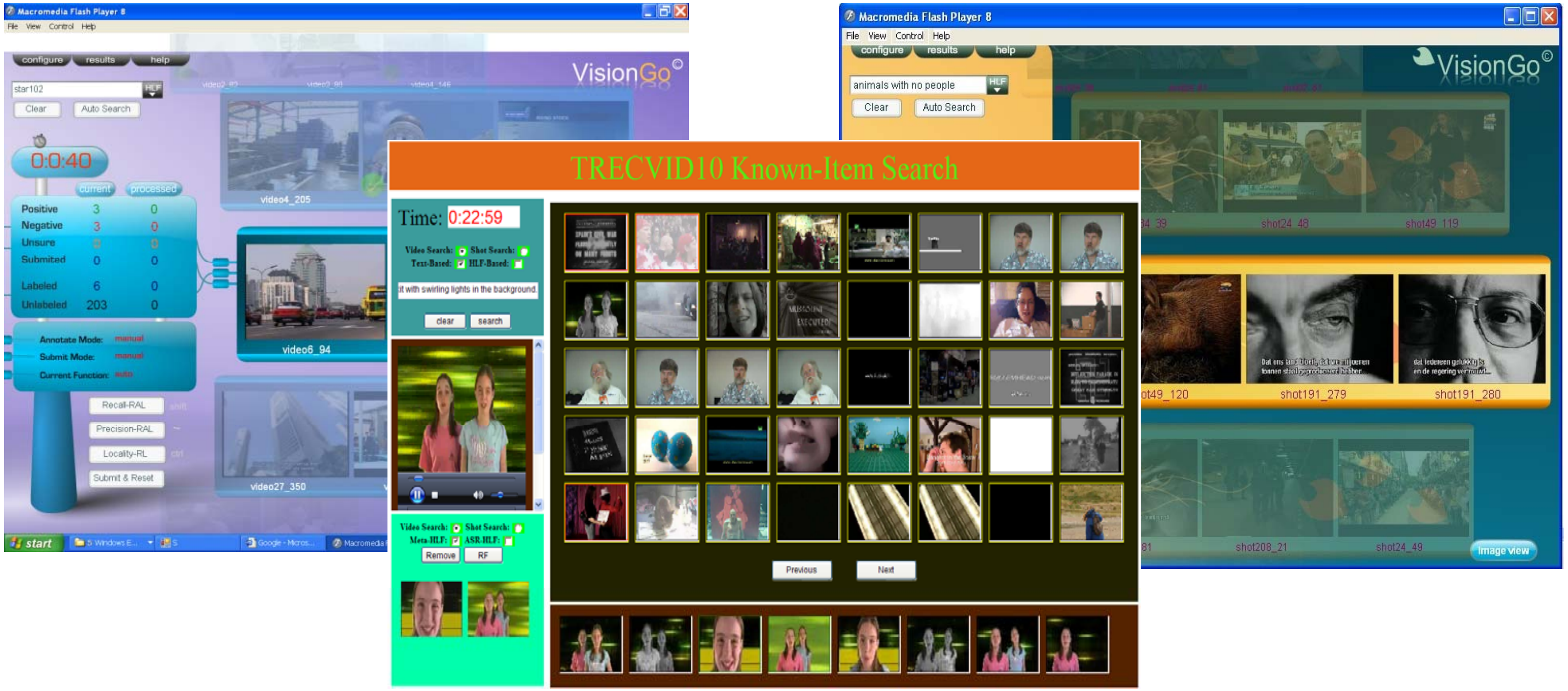
# Known-Item Search Task

- Given a text-only description of the video desired (Ground Truth Only One )
  - Automatically return a list of up to 100 video IDs ranked by probability. (5 minutes)
  - Interactively return the ID of the sought video and elapsed time to find it. (5 minutes)

0022 QUERY: Find the video of a man and woman getting dressed, a cat on window sill and another cat joining it, a wedding, two kittens and two babies

# Motivations

- Efficient web service oriented video interactive search
  - Efficient user interface (UI) for good interaction and efficient visualization
  - New feedback algorithm based on both related samples and exclusive negative samples;
  - Clustered shot-icons for fast previewing the main content of the videos.



# VisionGo System

## User Interface

- Maximize user's annotation effort
- Video-Show: rich visual and audio content
- Clustering based Shot-Icons: Top-rank Icon + Expand Icon

## Auto Search

- Multi-modality features fusion: Metadata, ASR, HLF and Youtube data
- Query Analysis

## Interactive Search

- Related samples strategy
- Exclusive negative sample selection
- Fusion of two kinds of HLF

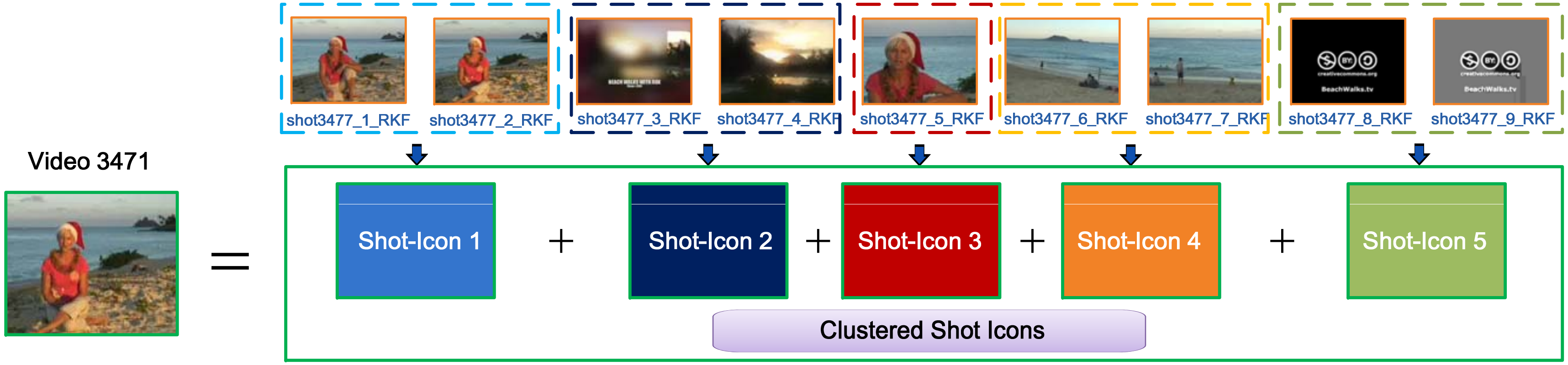
# Efficient User Interface

Maximize user's annotation effort

- Video-Show: show the detail and special visual and audio content

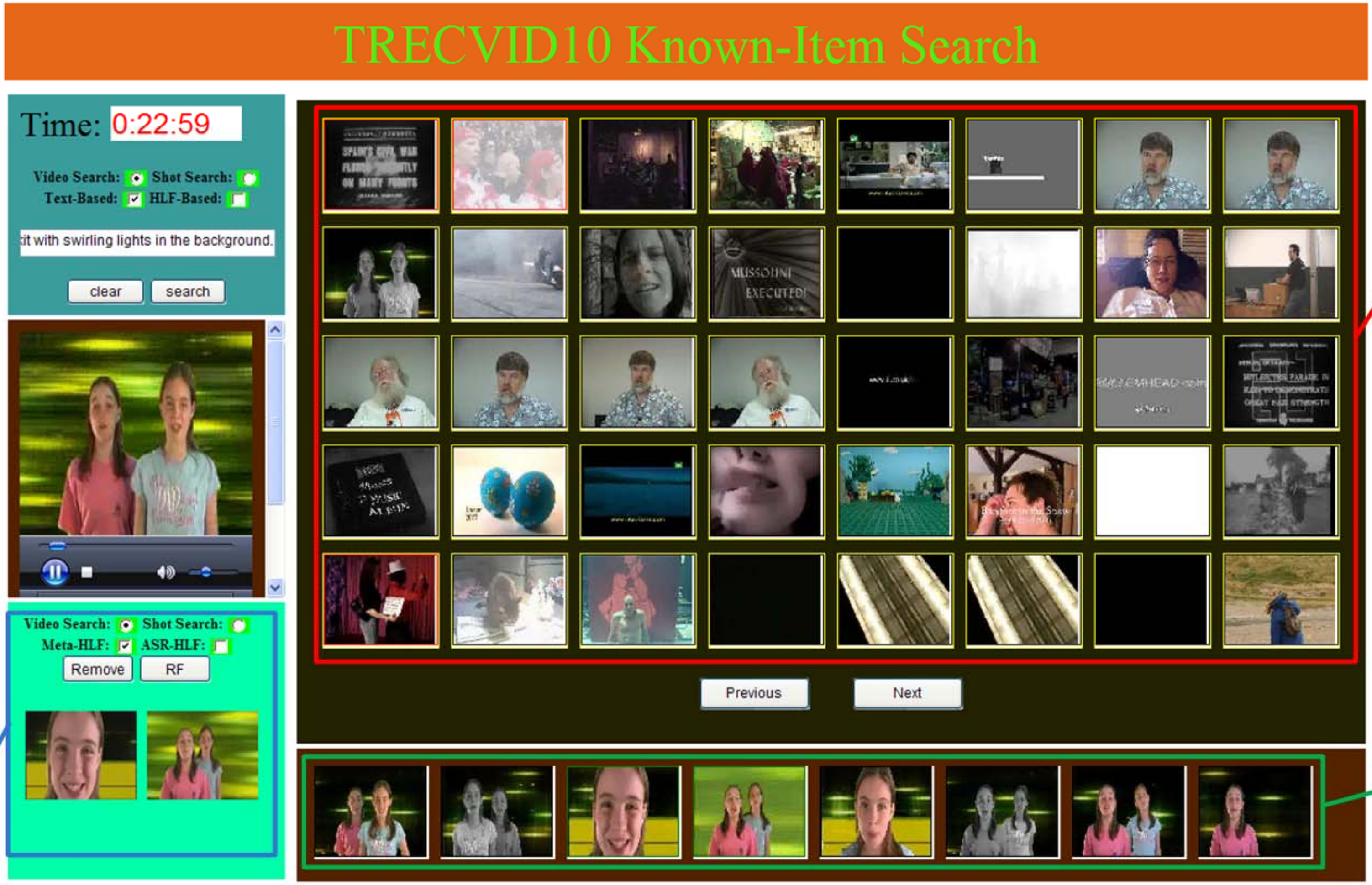
- Clustered Shot-Icons:

  Top-rank Icon + Expand Icon : represent the visual content of whole video



LMSearch

# Efficient User Interface



TRECVID10 Known-Item Search

Time: 0:22:59

Video Search: ● Shot Search: ○

Text-Based: ☑ HLF-Based: □

it with swirling lights in the background.

Video Search: ● Shot Search: ○

Meta-HLF: ☑ ASR-HLF: □

Results Show Area

Feedback Area

Related Sample Select Area

- ➢ **UI for good interaction and efficient visualization**
- ➢ **Maximize user’s annotation effort**

# Auto Search

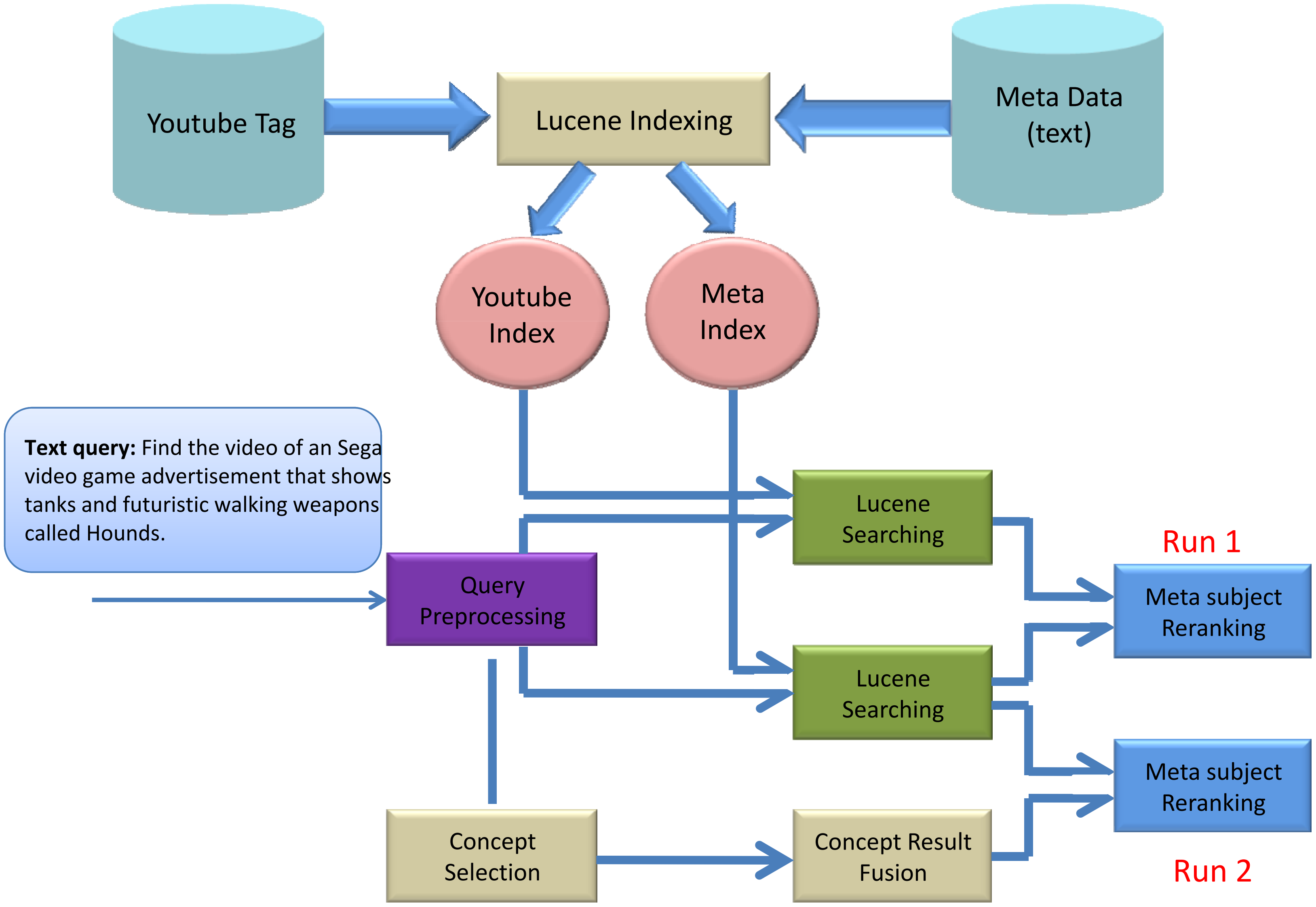
## Multi-modality features fusion

- Metadata is the most effective textual feature
- ASR plays a complementary role
- Tags of the crawled Youtube dataset

## Query Analysis

- Query expansion by Youtube
- Morphological analysis between description of HLFs and KIS’s queries

# Overview of Auto Search



**Text query:** Find the video of an Sega video game advertisement that shows tanks and futuristic walking weapons called Hounds.

# Query Analysis

- Query expansion by Youtube (two steps)

  (a) Use the query to retrieve relevant video from Youtube and collect the tags/comments
  (b) Extract terms from this collection (high mutual info.)

- Morphological analysis
  - HLF is necessary to query in terms of visual requirement
  - Utilize WordNet to do selective expansion
  - Match between feature descriptions of HLFs and KIS's queries

# Auto Search Performance

| Runs | Mean inverted rank | Mean elapsed time (mins) | Mean user satisfaction |
|---|---|---|---|
| Run1<br>(Metadata+<br>Youtube) | 0.215 | 0.021 | 6.0 |
| Run2<br>(Metadata+HLF) | 0.217 | 0.021 | 6.0 |

- Additional Tags data set is crawled from the Youtube website
- This dataset consists of 8,383 subsets of Youtube tags
- Each subset is downloaded corresponding to the title of each video

- **Tags in Youtube are diverse as the words in metadata**
- **Need further denoise  and extract key words in this dataset**

# Interactive Search

Related Sample Strategy

Exclusive Negative Samples Selection

Fusion of Two Kinds of HLF

# Related Sample Strategy

## Related Sample based Feedback

- Related sample refer to those video segments that are irrelevant to the query but relevant to some of the related concepts of the query. (Yuan el. CIVR10)
- New feedback strategy based on related shots of different videos

Shot query detector

$$f^t(x) = \eta^t \sum_{k=1}^{K} d_k^t f_k(x) + \frac{1}{t-1} \sum_{l=1}^{t-1} \beta_l^t \Delta f^l(x) + \Delta f^t(x)$$

Related Concept Detectors

Previous Delta Detector

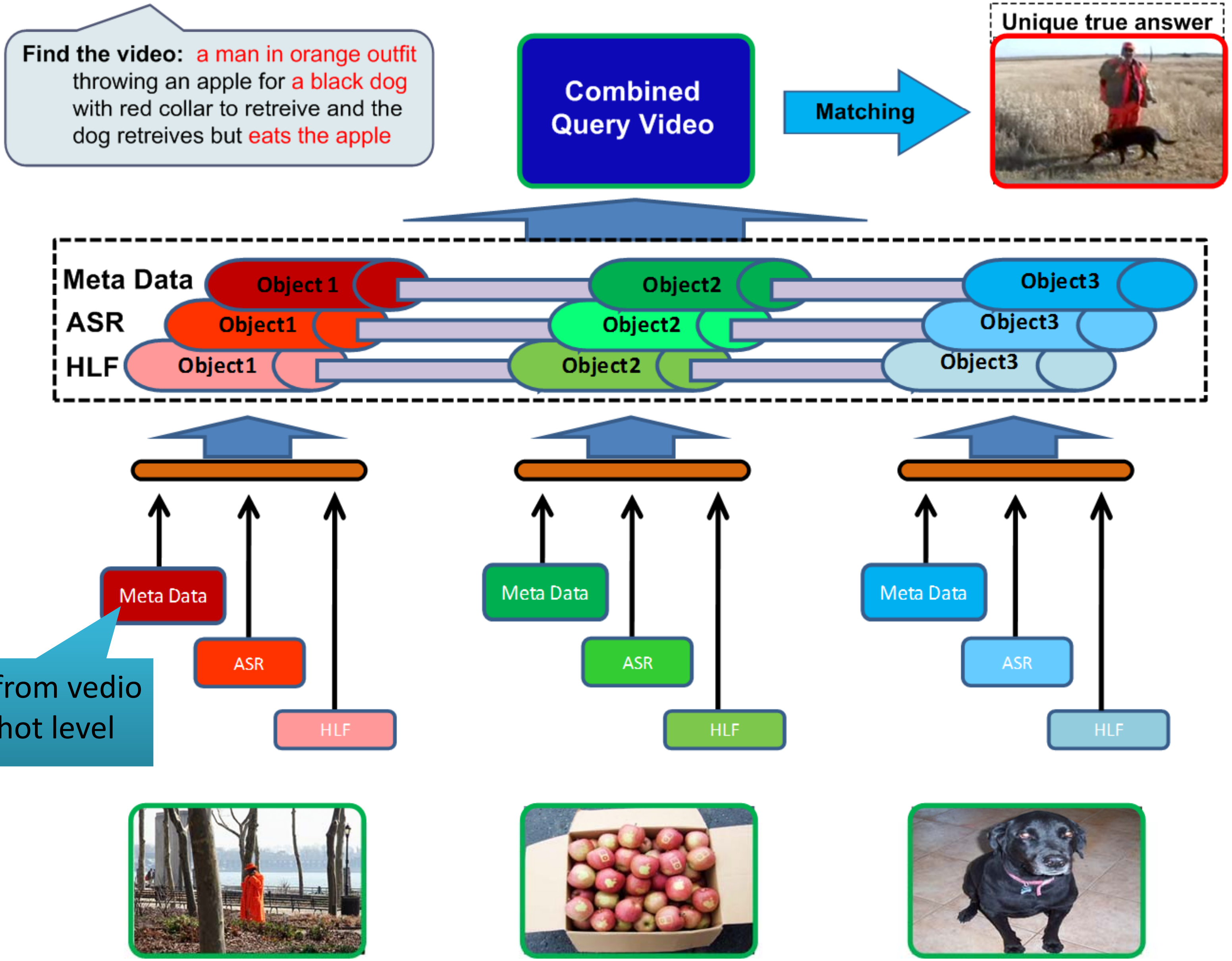
Current Delta Detector

### Learn Video Detector by Fusion

$$F^t(v_j) = \frac{1}{N_{v_j}} \sum_{p=1}^{N_{v_j}} f^t(s_p)$$

# Related Sample Strategy

Find the video: a man in orange outfit throwing an apple for a black dog with red collar to retreive and the dog retreives but eats the apple



Transfer from vedio level to shot level

# Exclusive Negative Samples Selection

## Exclusive Concept Subsets

$G_1$={airplane, infants, basketball, dancing, … , hospital, maps, laboratory }
$G_2$={telephones, birds, chair, basketball, … , flowers, golf, infants, maps}
$G_3$={laboratory, mountain, basketball, maps, … , singing, kitchen, driver}
……
$G_{n-1}$={golf, hospital, highway, infants, … , laboratory, prisoner, stadium}
$G_n$={boat_ship, cows, court, dancing, … , computer_or_televison_screen}

- **If the selected related samples contain the concepts: "birds", "mountain", "highway", then the exclusive negative set for the query is**

$$G_e = (G_2 \cup G_3 \cup G_{n-1}) \setminus \{\text{``}birds\text{''}, \text{``}mountain\text{''}, \text{``}highway\text{''}\}$$

- Construction for exclusive concept sets:

  *Robust Graph Mode Seeking by Graph Shift* (*Liu H. and Yan S. ICML'10* )

# Fusion of Two Kinds of HLF

- Linear Fusion Detector Scores (130 concepts):

  Multi-lable Propagation (Chen el. MM 2010) + CU-VIREO374 (Y.-G. Jiang el . 2008 )

- Visual features:

  225-D blockwise color moments
  128-D wavelet texture
  75-D edge direction histogram

- Advantages:

  - Computation cost: about 32 hours
  - Learned concept scores are robust to noises

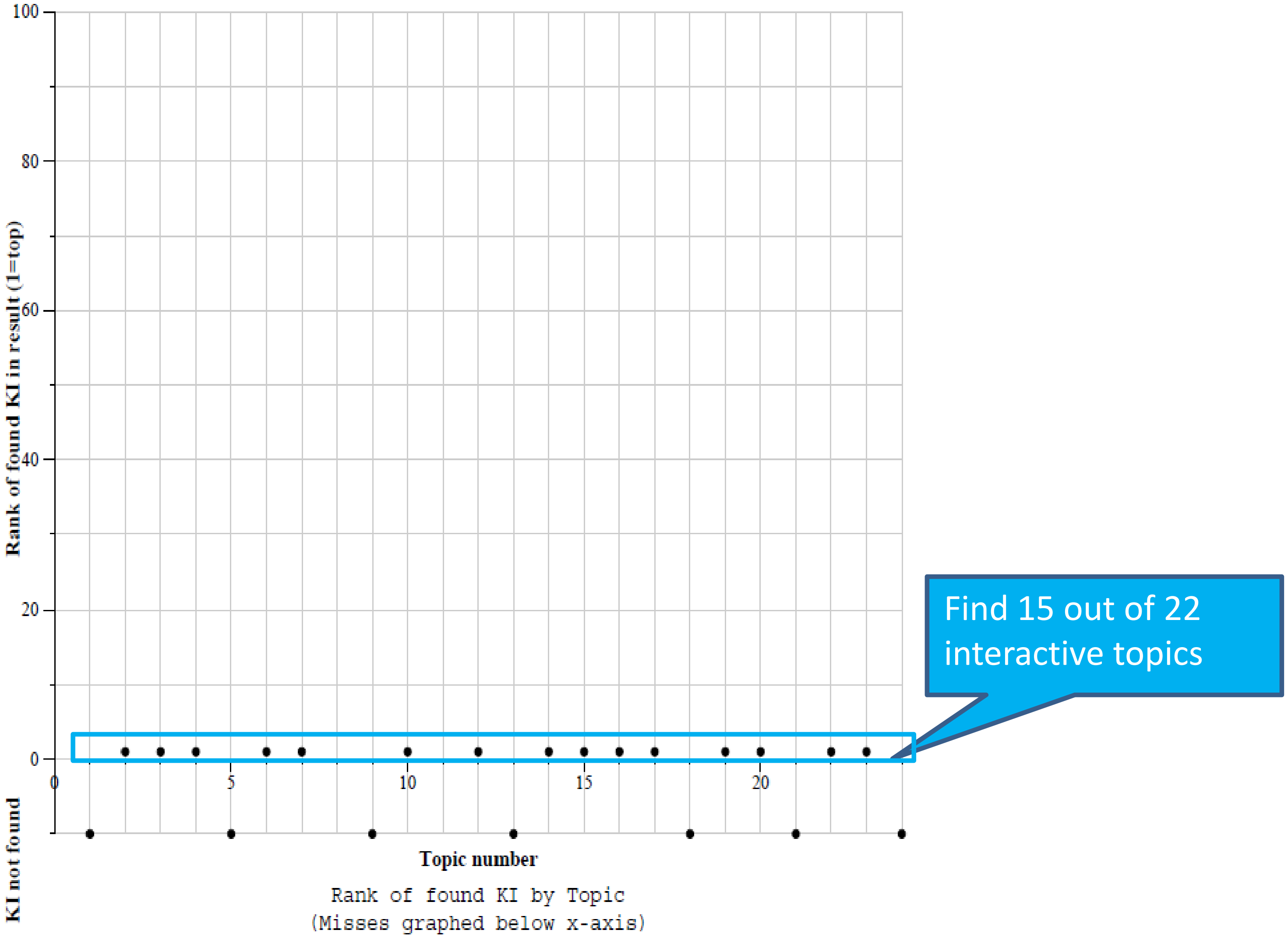
# Interactive Search Performance

| Runs | Mean inverted rank | Mean elapsed time (mins) | Mean user satisfaction |
|---|---|---|---|
| Run1 (Metadata+HLF) | 0.628 | 2.799 | 5.75 |
| Run2 (Youtube+HLF) | 0.628 | 2.577 | 6.0 |

➢ Top 2 performance in all interactive search participants

➢ Validate proposed feedback scheme based on both related samples and exclusive negative samples

# Interactive Search Performance



Rank of found KI by Topic
(Misses graphed below x-axis)

Find 15 out of 22 interactive topics

# Demo of VisionGo

Interactive QUERYs:

- Find the video of a man and women getting dressed, a cat on window sill and another cat joining it, a wedding, two kittens and two babies
- Find the video of one girl in a pink T shirt and another in a blue T shirt doing an Easter skit with swirling lights in the background
- Find the video of 21 seconds of your time featuring orange, Japanese lanterns in the night
- Find the video of the cost of drugs, featuring a man in glasses at a kitchen table, a video of Bush, and a sign saying Canada
- Find the video of President Bush standing near sea vessels with Coast Guard members talking about his pride of the Coast Guard, immigration, and security issues.
- Find the video of a street that has a pedestrian crosswalk indicated with blue stripes. People are walking on the sidewalk and cars are driving on the street

# Conclusions & Future Work

## Contributions in this work

- Efficient UI in interactive video search
- Proposed feedback method based on both related samples and exclusive negative samples
- Clustered shot icons for fast previewing main content of the videos

## Future work

- Extend the proposed novel feedback to real condition web services
- Develop more intuitive UI to enhance the user experience

Thank you!

NUS
National University
of Singapore

LMSearch