The Future of Multimedia and Video Retrieval

Myths and Opportunities

Alex Hauptmann
Dept. of Computer Science & Language Technologies Institute
Carnegie Mellon University,
Pittsburgh, PA, USA
alex@cs.cmu.edu
http://www.cs.cmu.edu/~alex

Video Analysis and Retrieval is Dead!

1. In the future, most metadata will be attached at creation time

All Metadata is Attached at Creation

- Cameras can record location, lighting, camera motion
- Editing actions will be remembered and connected to the video product
- Movies and sports events
  - High production value
  - High profit
  - High costs to create
- Incremental cost to do good manual annotation is marginal
- What about low value video production
  - YouTube, Flickr, etc. ?

Video Analysis and Retrieval is Dead!

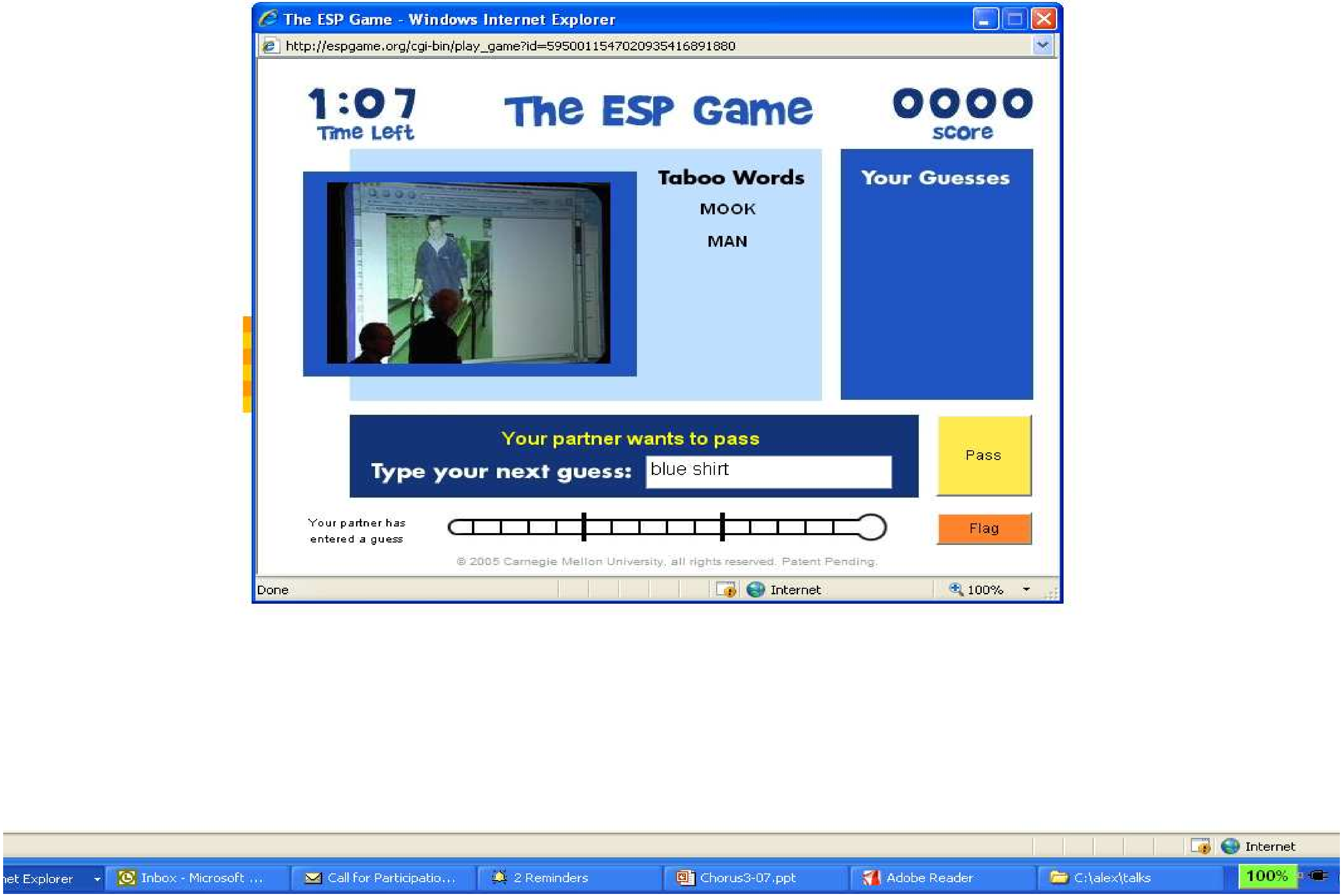
1. In the future, most metadata will be attached at creation time
2. Social video sharing sites can do any search much better than automated methods

Social Multimedia Sharing

Flickr, MySpace, YouTube, …

- User comments, annotations, tags, links
- Reasonable retrieval capability
- Everything will be done with social and human computation

Let's consider this



The ESP Game - Windows Internet Explorer

http://espgame.org/cgi-bin/play_game?id=5950011547020935416891880

1:07 Time Left

The ESP Game

0000 score

Taboo Words

MOOK

MAN

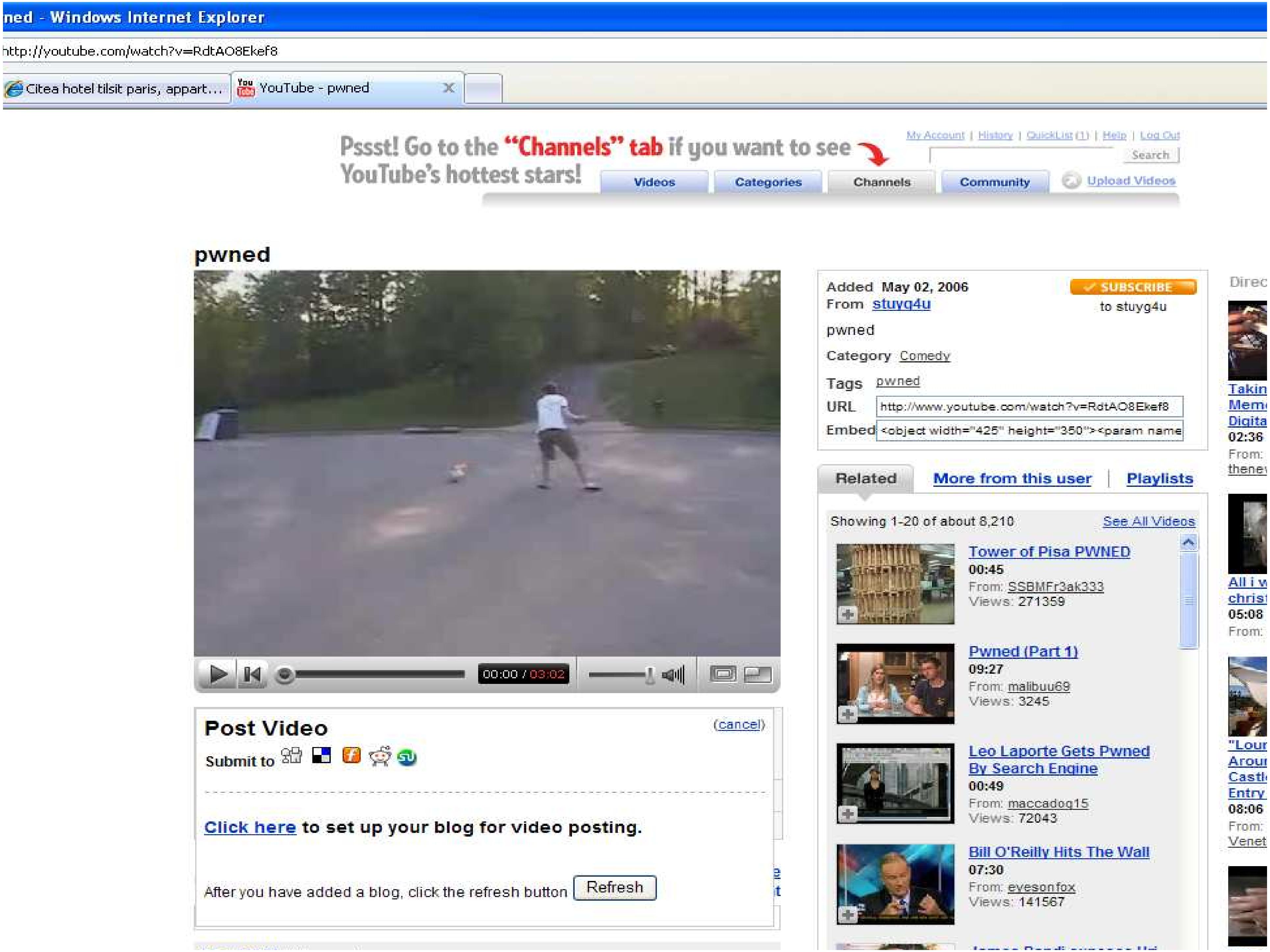
Your Guesses

Your partner wants to pass

Type your next guess: blue shirt

Your partner has entered a guess

© 2005 Carnegie Mellon University, all rights reserved. Patent Pending.



pwned

Added May 02, 2006
From stuyg4u

pwned

Category Comedy

Tags pwned

URL http://www.youtube.com/watch?v=RdtAO8Ekef8

Embed <object width="425" height="350"><param name

Showing 1-20 of about 8,210

Tower of Pisa PWNED
00:45
From: SSBMFr3ak333
Views: 271359

Pwned (Part 1)
09:27
From: malibuu69
Views: 3245

Leo Laporte Gets Pwned By Search Engine
00:49
From: maccadog15
Views: 72043

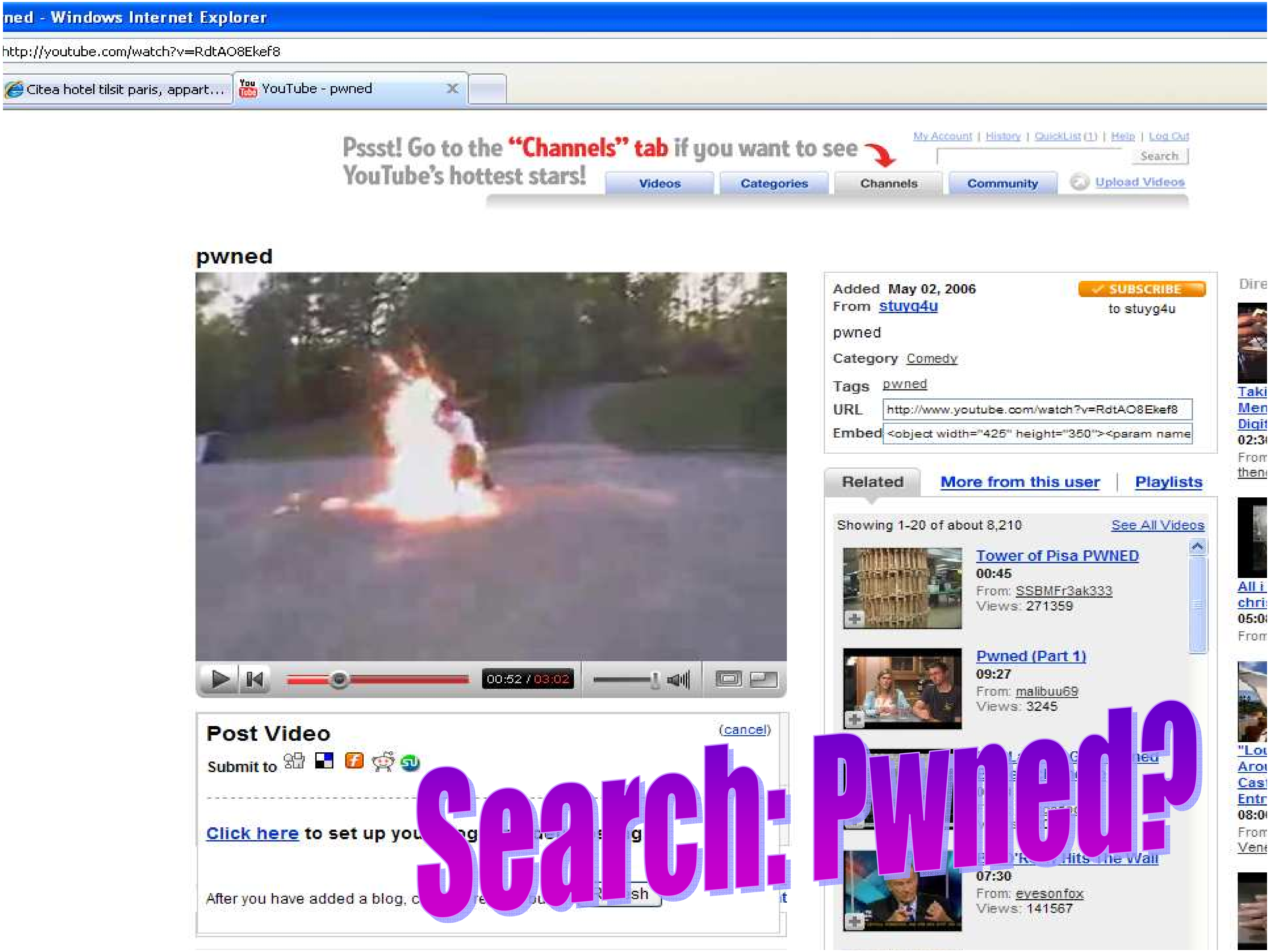
Bill O'Reilly Hits The Wall
07:30
From: eyesonfox
Views: 141567

Post Video (cancel)

Submit to

Click here to set up your blog for video posting.

After you have added a blog, click the refresh button Refresh



ned - Windows Internet Explorer

http://youtube.com/watch?v=RdtAO8Ekef8

Citea hotel tilsit paris, appart... | YouTube - pwned

Pssst! Go to the "Channels" tab if you want to see YouTube's hottest stars!

pwned

Added May 02, 2006
From stuyg4u

pwned

Category Comedy

Tags pwned

URL http://www.youtube.com/watch?v=RdtAO8Ekef8

Embed <object width="425" height="350"><param name

Showing 1-20 of about 8,210 See All Videos

Tower of Pisa PWNED
00:45
From: SSBMFr3ak333
Views: 271359

Pwned (Part 1)
09:27
From: malibuu69
Views: 3245

07:30
From: eyesonfox
Views: 141567

Post Video (cancel)

Submit to

Click here to set up you

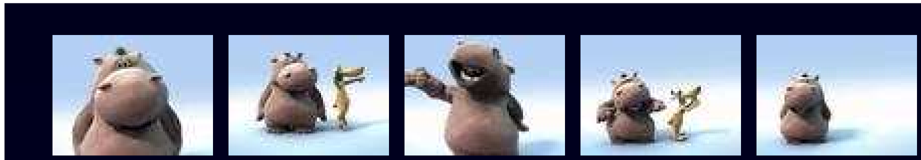
After you have added a blog, c

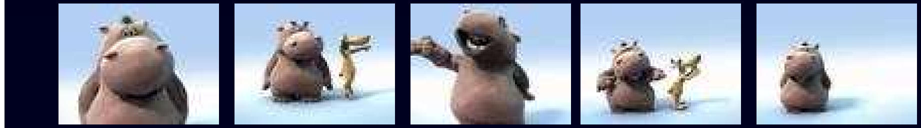
Search: Pwned?

Query

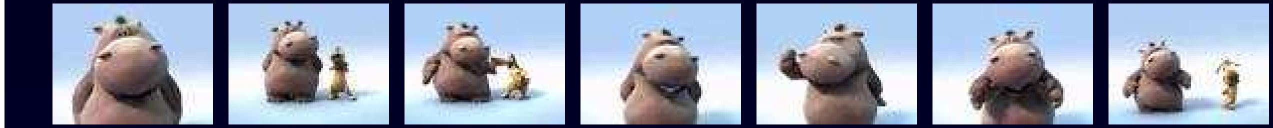
Distance

“The lion sleeps tonight”





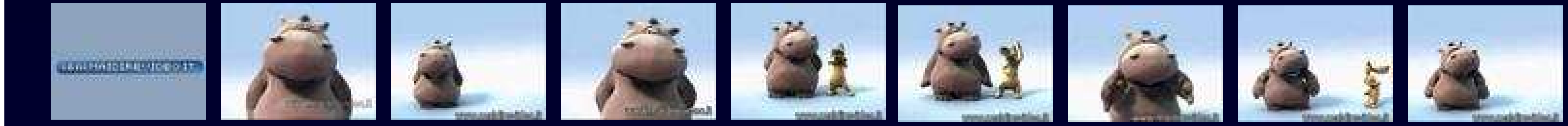
0



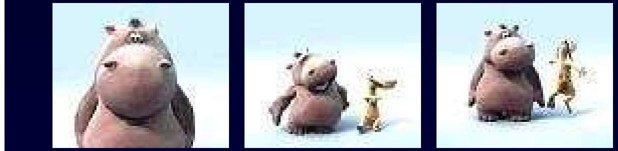
0.01

UNREGISTERED VERSION OF SNAPZ PRO
BLACK PEOPLE WILL PREVAIL IN THE WORLD

0.91



0.33

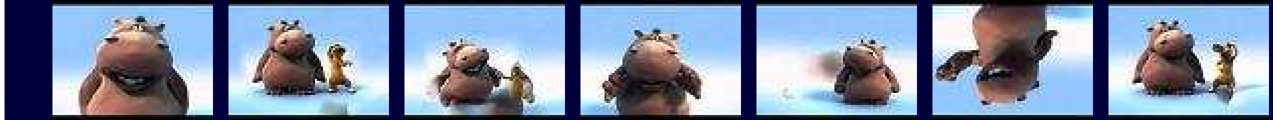


0.64

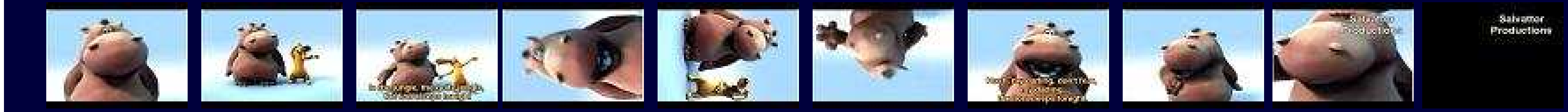
jetzt wird's LUSTIG.DE
BILLIGREISEN.DE

0.76

0.67



0.46

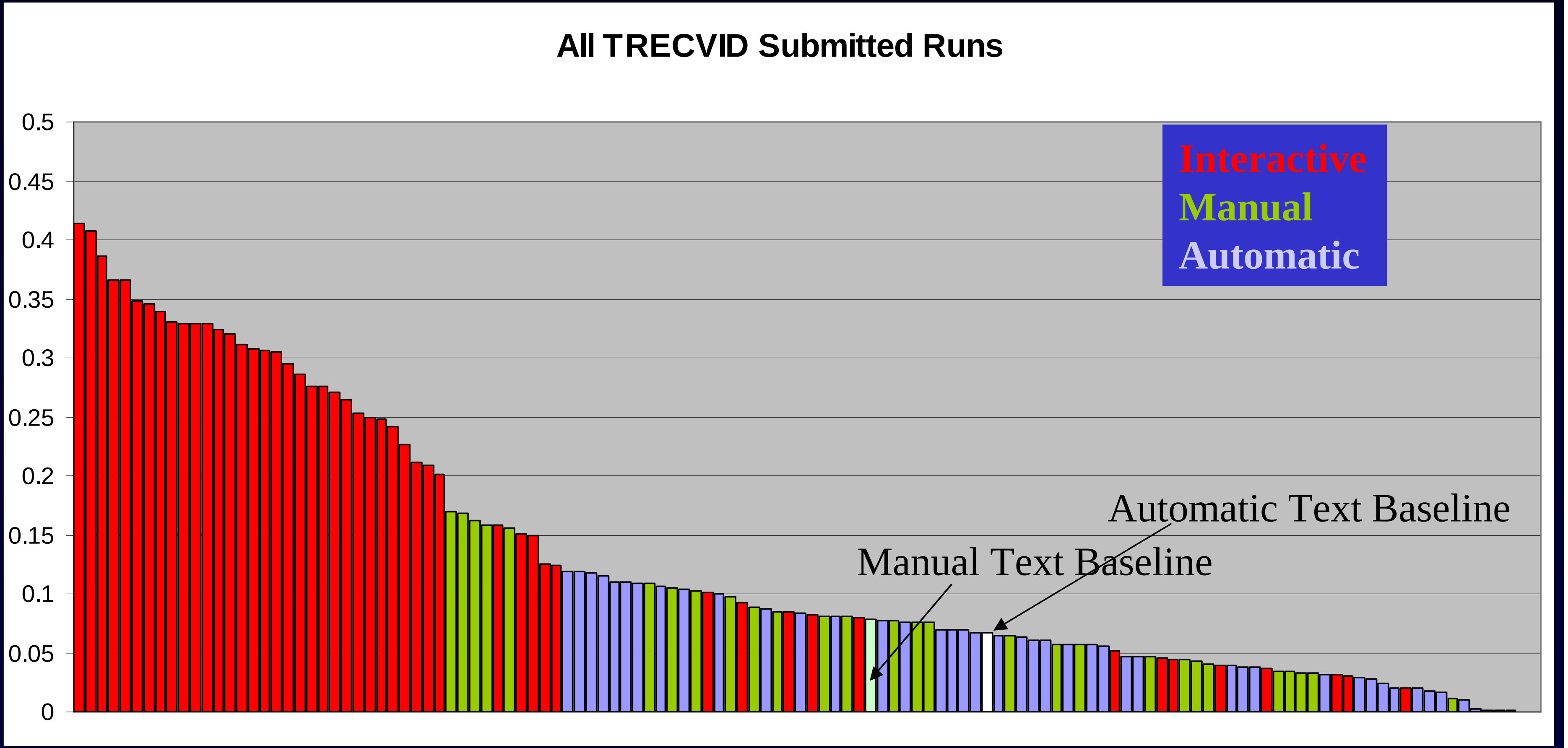


0.51

Video Analysis and Retrieval is Dead!

1. In the future, most metadata will be attached at creation time
2. Social video sharing sites can do any search much better than automated methods
3. Video retrieval doesn’t work any better than text search
   - TrecVid 2001 - 2003

TRECVID 2005 System Comparisons



Differences between best systems and baselines are significant

Accuracy for non-interactive systems is consistently **LOW**

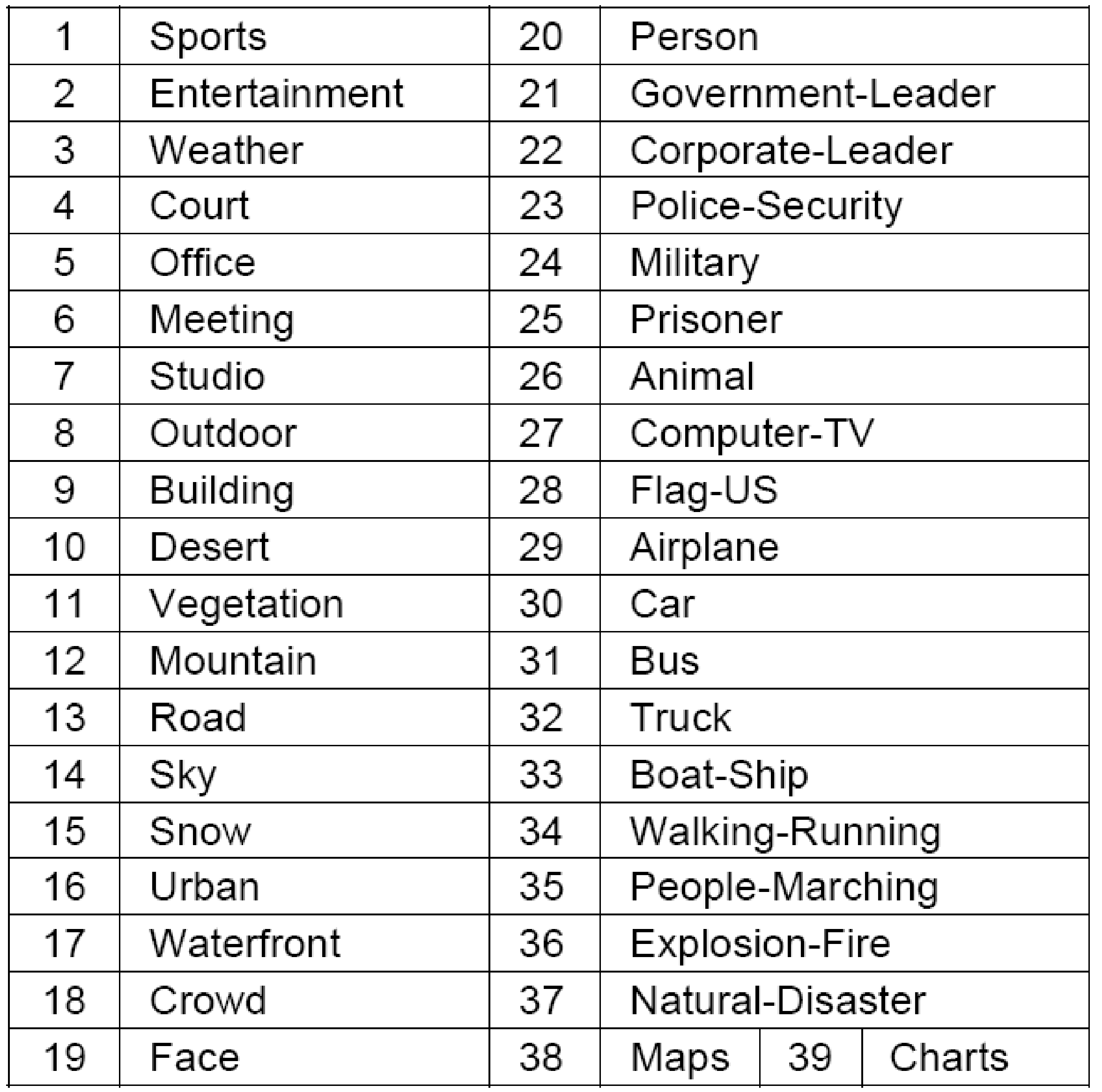
What Makes Video Retrieval Work?

- Low level visual features are not sufficient to understand an image or video clip ("Semantic Gap")
  - Low-level: Texture, color, shape, interest points, motion, audio (SFFT, MelCep, Zero crossing, …)
- Describe video through intermediate ***semantic*** concepts
  - Face, car, outdoors, boat, building, clouds, sky, water, …
- Semantic concepts can be learned automatically
- Semantic concepts are useful for retrieval

Why are Semantic Concepts Important?

- What if we could detect a lot of concepts?
- Speech recognition analogy
  - ~~100 words~~ → ~~1000 words~~ → 20,000 words → 64,000 words
- LSCOM – A Large Scale Ontology for Multimedia
  - 2 year workshop to define and annotate 1000 concepts
  - Defined 850 concepts
  - Extended via ontology to ~2400 concepts,
  - Annotated 450 concepts on 70 hours of TV news
  - Available at www.LSCOM.org

39 Semantic Concepts (LSCOM-Lite)



1	Sports	20	Person		
2	Entertainment	21	Government-Leader		
3	Weather	22	Corporate-Leader		
4	Court	23	Police-Security		
5	Office	24	Military		
6	Meeting	25	Prisoner		
7	Studio	26	Animal		
8	Outdoor	27	Computer-TV		
9	Building	28	Flag-US		
10	Desert	29	Airplane		
11	Vegetation	30	Car		
12	Mountain	31	Bus		
13	Road	32	Truck		
14	Sky	33	Boat-Ship		
15	Snow	34	Walking-Running		
16	Urban	35	People-Marching		
17	Waterfront	36	Explosion-Fire		
18	Crowd	37	Natural-Disaster		
19	Face	38	Maps	39	Charts

Annotated Concept Sets

- Trecvid 2006 development data
  - ~70hours English, Arabic, Chinese News
  - 62000 shots

3 Annotated Concept Sets:

- LSCOM Lite
  - 39 concepts
- Media Mill
  - 75 concepts that overlap with LSCOM
- LSCOM
  - 300 concepts
  - Minimal frequency cutoff

Speculative Scenario with Lots of Concepts

[Hauptmann et al, CIVR2007]

Best Case:

- Perfect concept detection (Oracle)
- Perfect concept combination (Oracle)

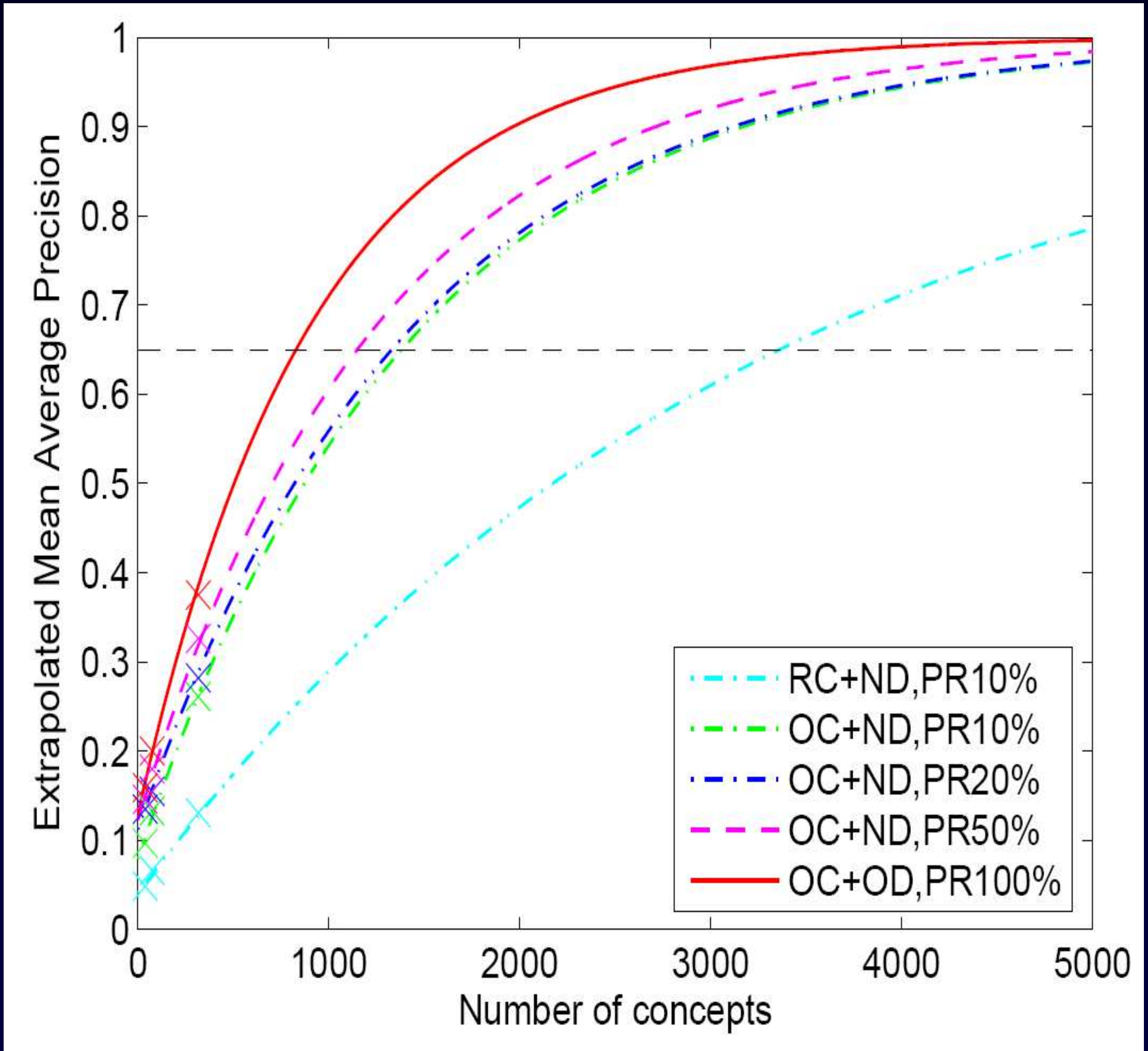
- Noisy detection (different error rates)
- Realistic combination (50%)

Extrapolation Assumption:

- Things get harder as you add more concepts
  - Proportional to the difference between the current MAP and the upper limit of 1
    I.e. the higher the current MAP, the less benefit a new concept offers

How well can you retrieve relevant shots (documents)

Extrapolation Results



Conjecture: ~4000 concepts are enough

[CIVR2007]

Opportunities in Multimedia Retrieval Beyond the Standard Paradigm

Retrieval with robust semantic concepts

- Ontology?

Retrieval of web video

- Duplicate removal
- Summarization and preview
- Combine social network analysis and content analysis

Retrieval from long-term surveillance

- No human annotation possible
- Collaboration with of multimedia, computer vision and information retrieval
- Nursing home Scenario

CareMedia:
Automated Behavior Analysis in the Nursing Home

Longitudinal video and sensor analysis into semantic concepts

- Automating detection of behavioral & psychological symptoms of dementia

Goal: Monitoring and maintaining the quality of life

Automated, quantitative measurements to:

- Explore relationship of dementia to environments in which they occur
- Evaluate symptoms longitudinally
- Determine of the frequency of symptoms
- Develop a patient profile of responses to pharmacological and non-pharmacological interventions

>>>> Enable earlier intervention to sustain quality of life

CareMedia: What are the observables?

- Who?
  - Identify people across cameras, days
- What are they doing?
  - Wandering around
  - Working on tasks
  - Looking for things
  - Eating, sleeping
- How well did they do it?
  - Quantify performance
  - Detect/report anomalies

Opportunities in Multimedia Retrieval Beyond the Standard Paradigm

Retrieval with robust semantic concepts

Ontology?

Retrieval of web video

- Duplicate removal
- Summarization and preview
- Combine social network analysis and content analysis

Retrieval from long-term surveillance

- No human annotation possible
- Collaboration with of multimedia, computer vision and information retrieval
- Nursing Home Example

- Integrate retrieval from sensors with video, audio and text data
  - Digital Human Memory example

Digital Human Memory

- Technology for creating a continuously recorded, digital, high fidelity record of one’s whole life in video form
- Personal, mobile units which record audio, video, GPS and electronic communications (wifi, bluetooth), body sensor data; capturing all that is heard, seen & experienced
- Transforming this personal history into a meaningful, accessible information resource

- Feasible: ~200MB/h or 2GB/day or .66 TB/year or 60 TB/lifetime

Opportunities in Multimedia Retrieval Beyond the Standard Paradigm

Retrieval with robust semantic concepts

Ontology?

Retrieval of web video

- Duplicate removal
- Summarization and preview
- Combine social network analysis and content analysis

Retrieval from long-term surveillance

- No human annotation possible
- Collaboration with of multimedia, computer vision and information retrieval
- Nursing Home Example

- Integrate retrieval from sensors with video, audio and text data
  - Digital Human Memory example

Less studied areas:

- Analysis of emotion in video
- Analysis of bias and perspectives in editing and presentation
- Insert advertising into video
- Tools for video creation and video mashups

New paradigms for information access for imperfect data

Thank You