



Utilizing Semantic Word Similarity Measures for Video Retrieval



Yusuf Aytar
Computer Vision Lab,
University of Central Florida
yaytar@cs.ucf.edu, yusufaytar@google.com

Mubarak Shah
Computer Vision Lab,
University of Central Florida
shah@cs.ucf.edu

Jiebo Luo
R&D Laboratories,
Eastman Kodak Company
jiebo.luo@kodak.com

Problem Overview

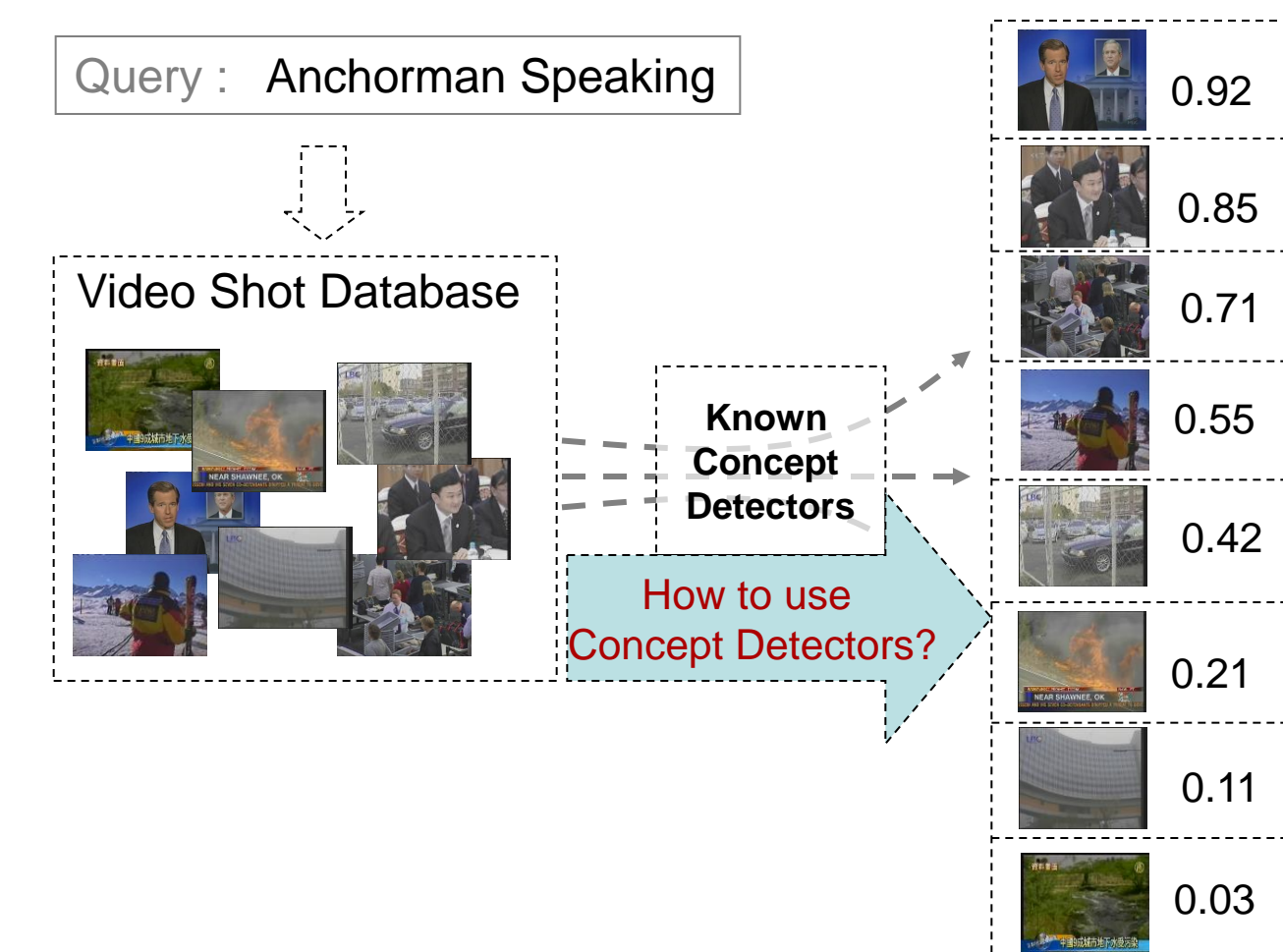
Motivation

- It's not possible to train detectors for all of the concepts in the real world.
- Available concept detectors can be used for retrieving **new concepts**.

Goal

- Retrieving **new concepts** with the help of available (**known**) **concept** detectors and semantic word similarity measures.

Problem Definition



Similarity Measures

1. Visual Co-occurrence

The co-occurrence of concepts in the same scene.

$$p(c_{Car}) = 0.067$$

$$p(c_{Truck}) = 0.011$$

$$p(c_{Car} \& c_{Truck}) = 0.0072$$

$$PMI_{Visual}(c_{Car}, c_{Truck}) = \log\left(\frac{p(c_{Car} \& c_{Truck})}{p(c_{Car})p(c_{Truck})}\right) = 2.3244$$

$$Sim_{Visual}(c_{Car}, c_{Truck}) = Sigmoid(PMI_{Visual}(c_{Car}, c_{Truck})) = 0.91$$

2. Semantic Word Similarity

Semantic word similarity is the relatedness of two concepts and it's generally a common sense knowledge that we build for years.

2.1. PMI-IR Similarity

Pointwise mutual information using data collected by information retrieval [Turney'01].

$$p(c_{Car}) = 0.097$$

$$p(c_{Truck}) = 0.020$$

$$p(c_{Car} \& c_{Truck}) = 0.0049$$

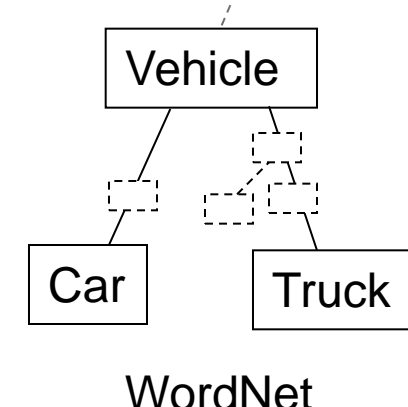
$$PMI_{IR}(c_{Car}, c_{Truck}) = \log\left(\frac{p(c_{Car} \& c_{Truck})}{p(c_{Car})p(c_{Truck})}\right) = \log\left(\frac{hits(c_{Car}, NEAR(c_{Truck})) \times WebSize}{hits(c_{Car})hits(c_{Truck})}\right) = 0.943$$

$$Sim_{PMI-IR}(c_{Car}, c_{Truck}) = Sigmoid(PMI_{IR}(c_{Car}, c_{Truck})) = 0.72$$

2.2. Lin's Similarity Measure

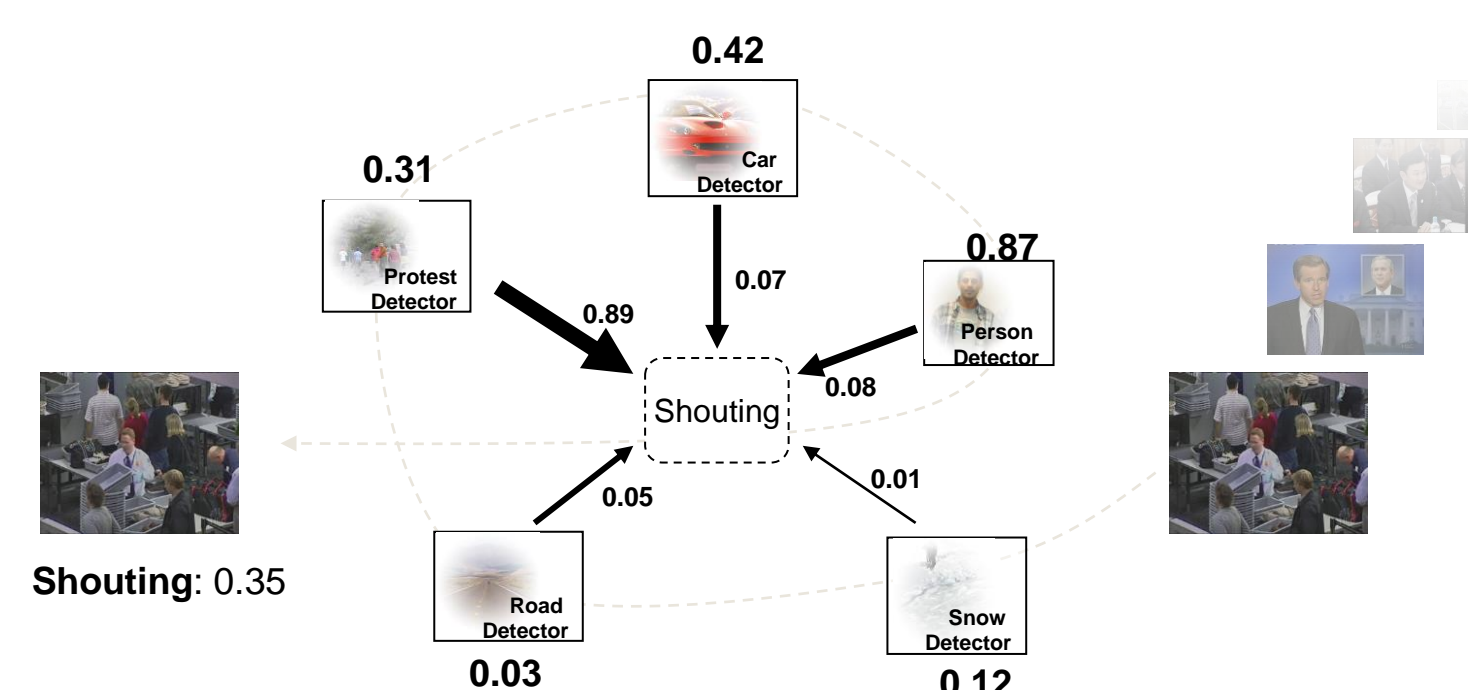
The key idea is to find the maximum information shared by two concepts and normalize it [Lin'98].

$$Sim_{Lin}(c_{Car}, c_{Truck}) = \frac{2 \times IC(LCS(c_{Car}, c_{Truck}))}{IC(c_{Car}) + IC(c_{Truck})} = 0.78$$

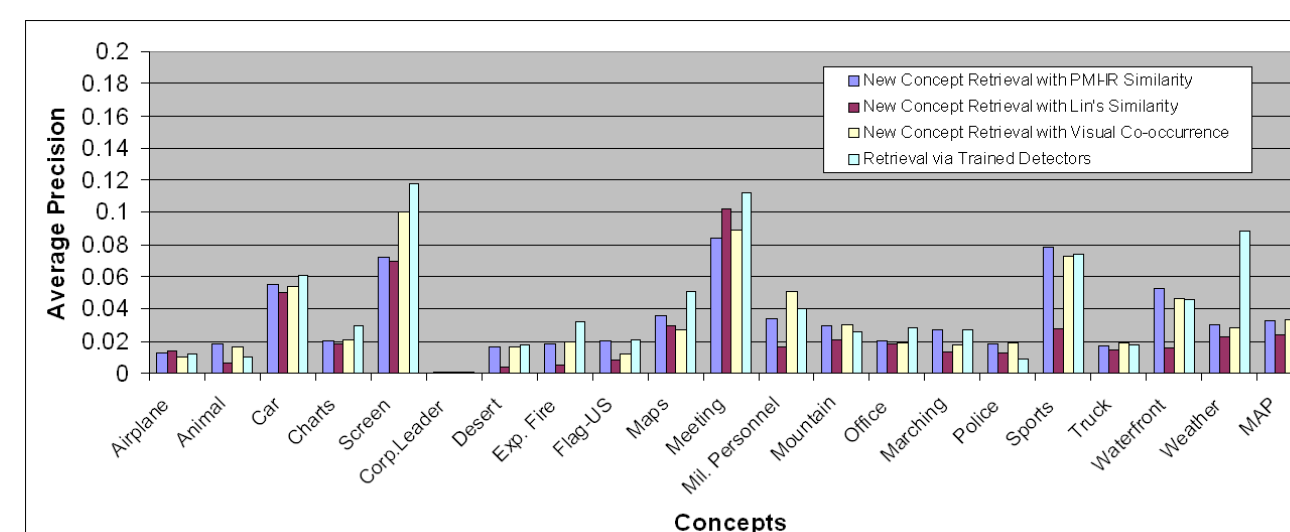


Retrieving New Concepts

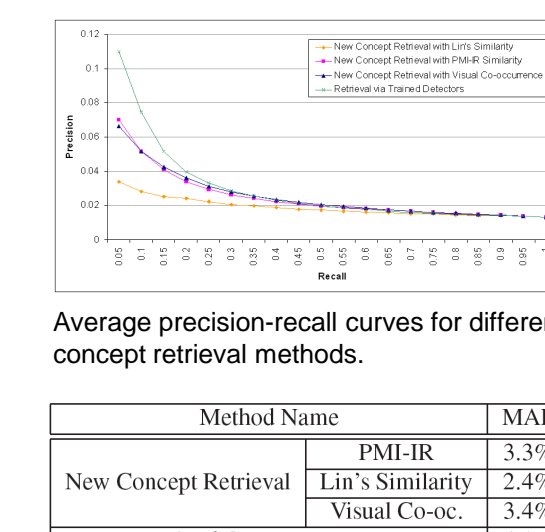
The score for the **new concept** is the linear combination of scores of **known concepts** and similarities between new concept and known concepts.

$$Score_{c_n}(s_k) = \frac{\sum_{j=1}^M Sim(c_j, c_n) Score_{c_j}(s_k)}{\sum_{j=1}^M Score_{c_j}(s_k)}$$


Results



A comparison of different retrieval methods using Average Precision. MAP (Mean Average Precision) is shown at far right.



MAP comparison of different concept retrieval methods.

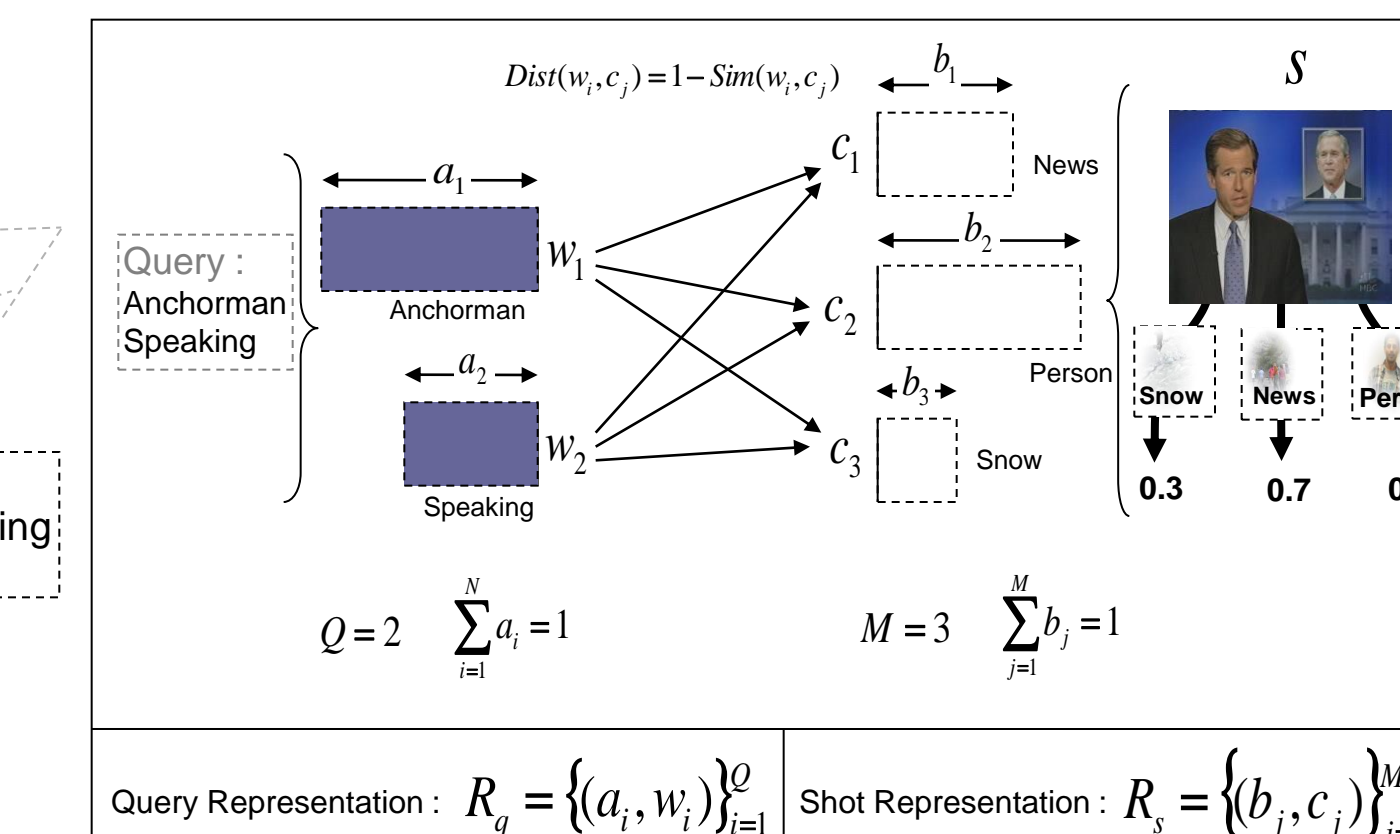


Top 100 retrieval results for the new concept "shouting".



Top 100 retrieval results for the new concept "fishing".

Semantic Retrieval

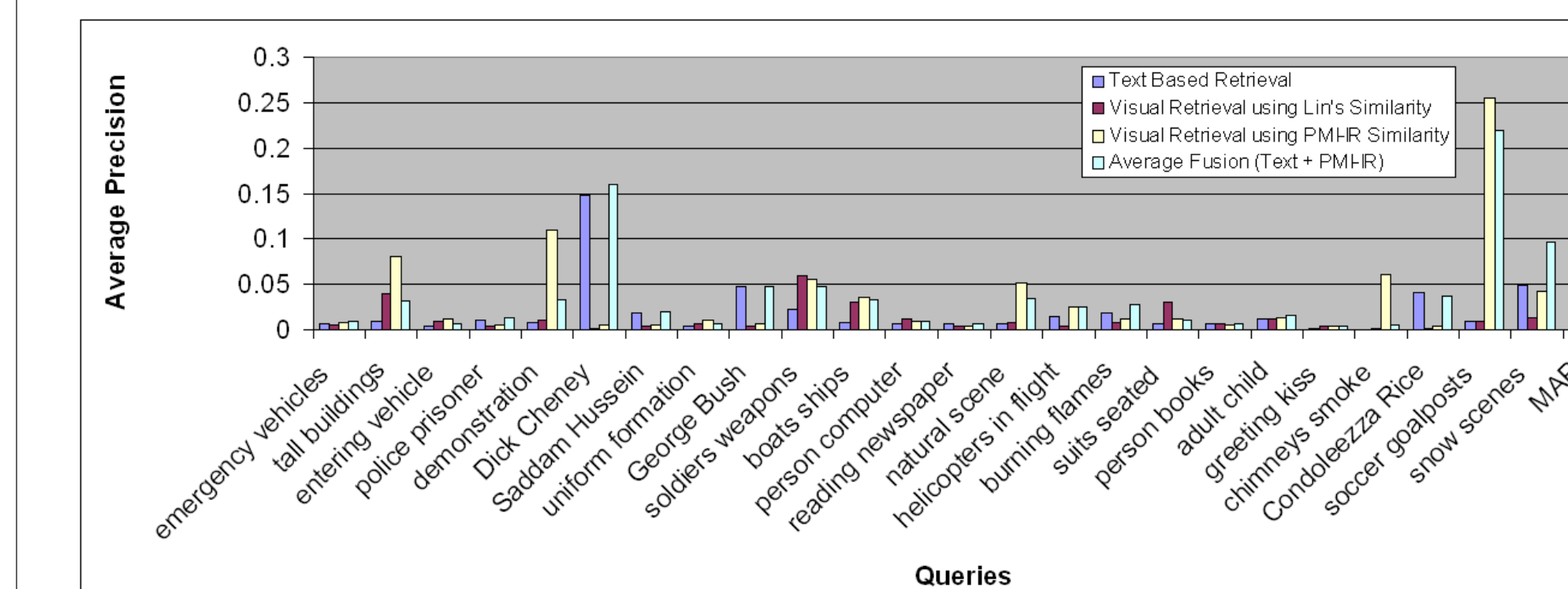


$$R_q = \{(a_i, w_i)\}_{i=1}^Q \rightarrow R_s = \{(b_j, c_j)\}_{j=1}^M$$

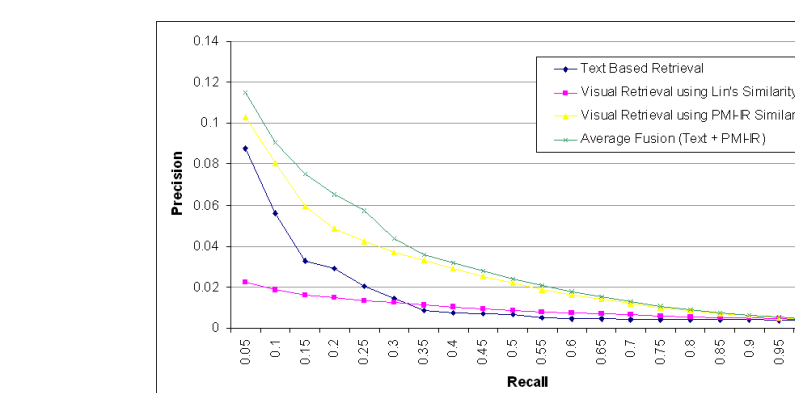
$$EMD(R_q, R_s) = \min_{F=\{f_{ij}\}} \sum_{i,j} f_{ij} Dist(w_i, c_j)$$

$$Score_q(s_k) = 1 - EMD_{s_k}(R_q, R_s)$$

Results



AP (Average Precision) results for different semantic retrieval methods for each query (represented by some selected words for each query) in TRECVID'06. MAP (Mean Average Precision) is shown at far right.



Average precision-recall curves for different semantic retrieval methods on TRECVID'06 queries.

Semantic Retrieval Method	MAP'06	MAP'07
Vision Based	3.4%	3.6%
Retrieval	1.1%	2.1%
Text Based Retrieval	1.9%	1.6%
Average Fusion (Text + PMI-IR)	3.7%	3.5%

MAP comparison for different semantic retrieval methods on TRECVID'06 and TRECVID'07 test data sets.

This research is partially funded by the U.S. Government's VACE program.

