IBM Research TRECVID 2005 Automatic Search System

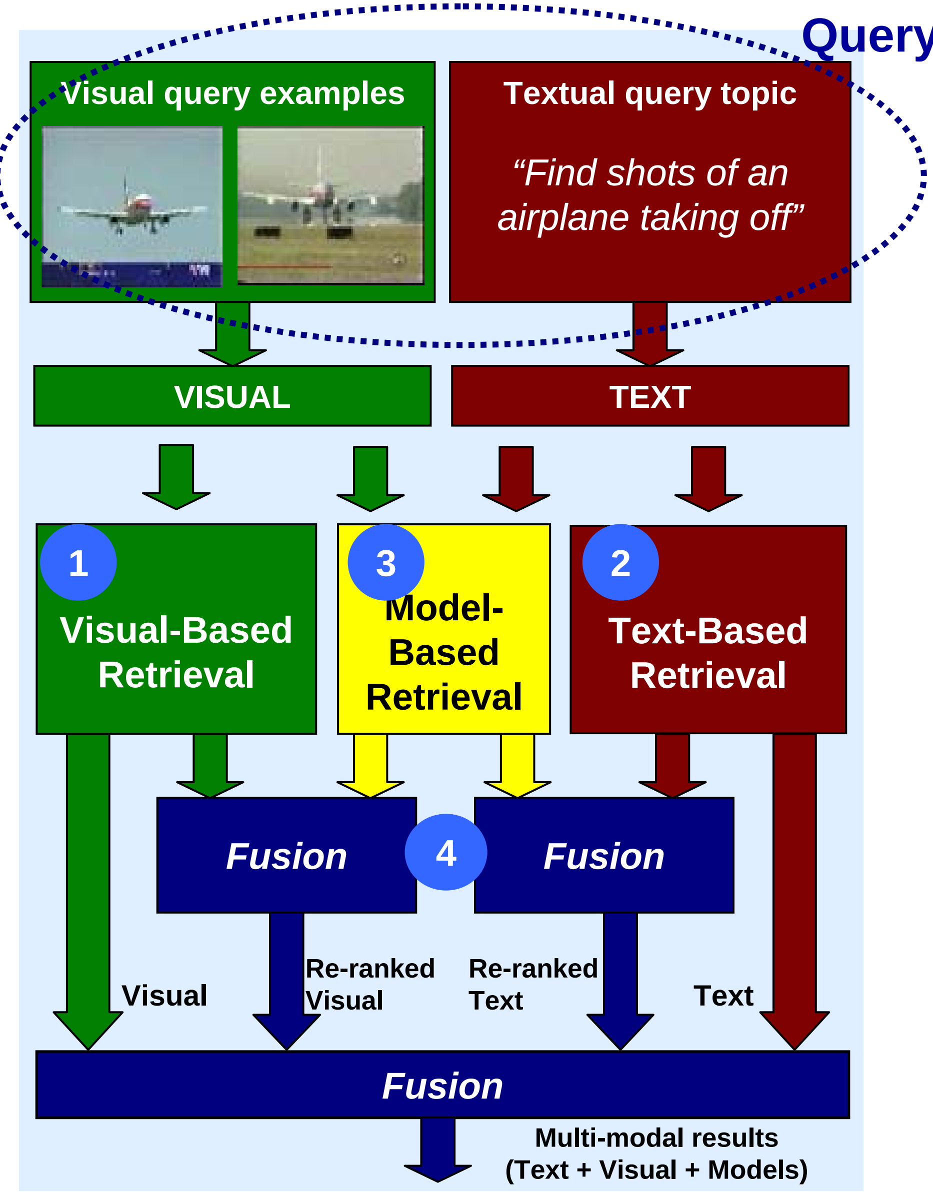
M. Campbell, S. Ebadollahi, A. Haubold, M. Naphade,
A. Natsev, J. R. Smith, J. Tešić, T. Volkmer
Intelligent Information Management Dept.
IBM Thomas J. Watson Research Center

Presented by Apostol (Paul) Natsev, natsev@us.ibm.com

NIST TRECVID Workshop
Gaithersburg, MD
November 14-15, 2005

ARDA
Video Analysis
Content Extraction

IBM Research Automatic Search System Overview



Approaches:

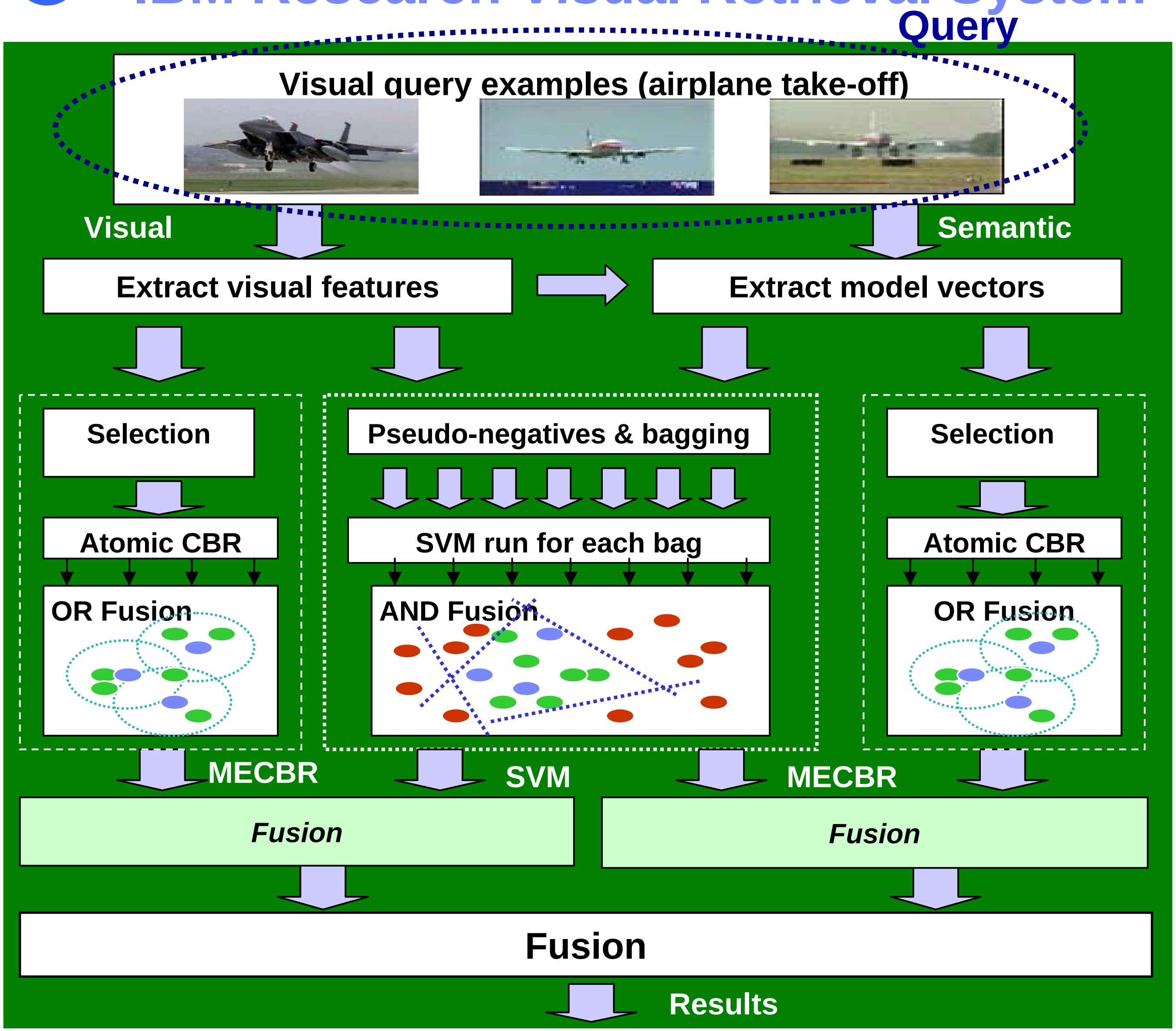
1. Visual-based: light-weight learning (discriminative and nearest neighbor modeling)
2. Text- based: automatic query expansion
3. Model-based: automatic query-to-model mapping & weighting
4. Fusion:
   - Query-independent
   - Statistical normalization (visual)
   - Rank normalization (text)
   - Model-based re-ranking (text & visual)

Highlights:

- Highest MAP for automatic type "A"
- Automatic search outperformed 24 of 26 manual search submissions

1

IBM Research Visual Retrieval System



Approaches:

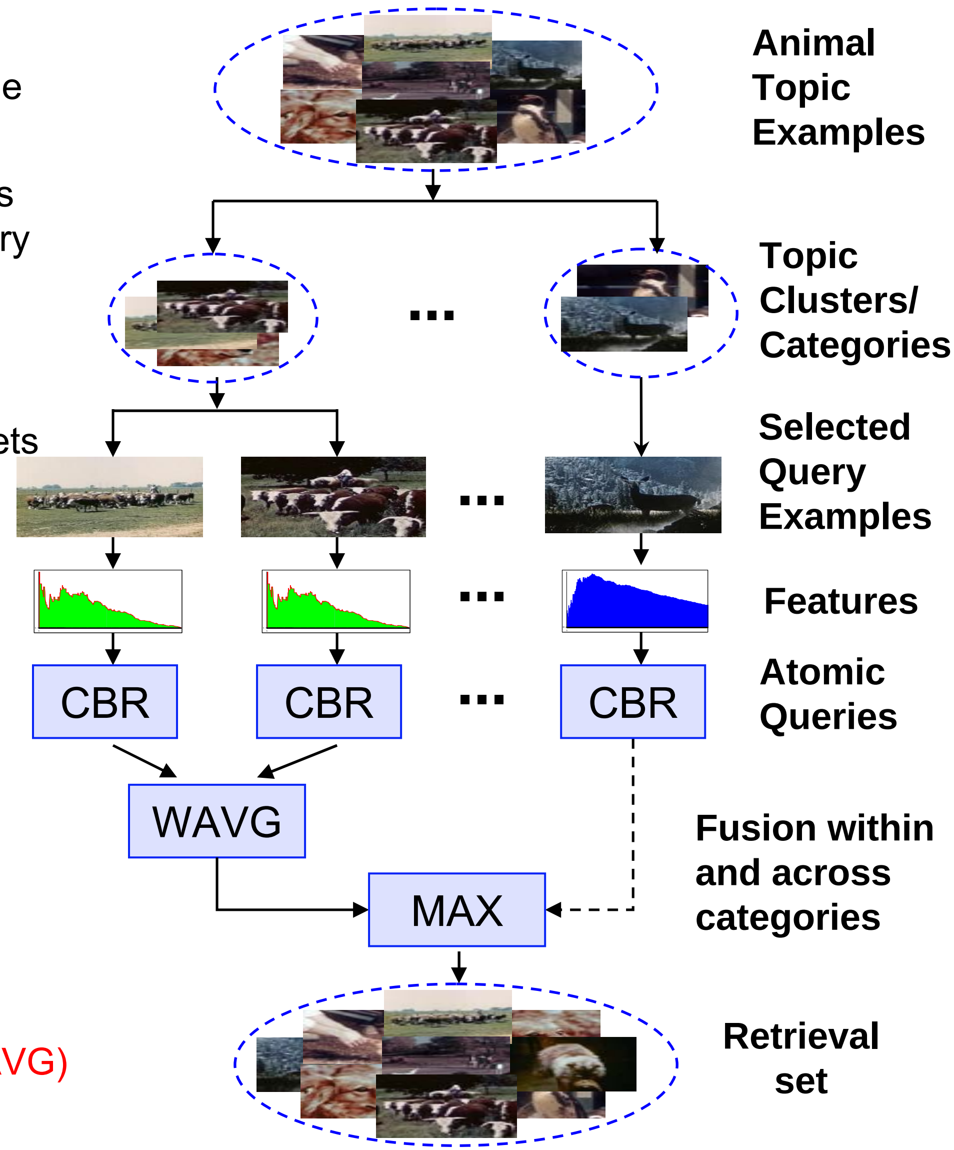
- Fusion of two light-weight learning techniques:
  - •MECBR (k-NN)
  - •SVM
- Low-level and high-level features

Highlights:

- Outperformed text- and model-based automatic type-“A” approaches at TRECVID-2005

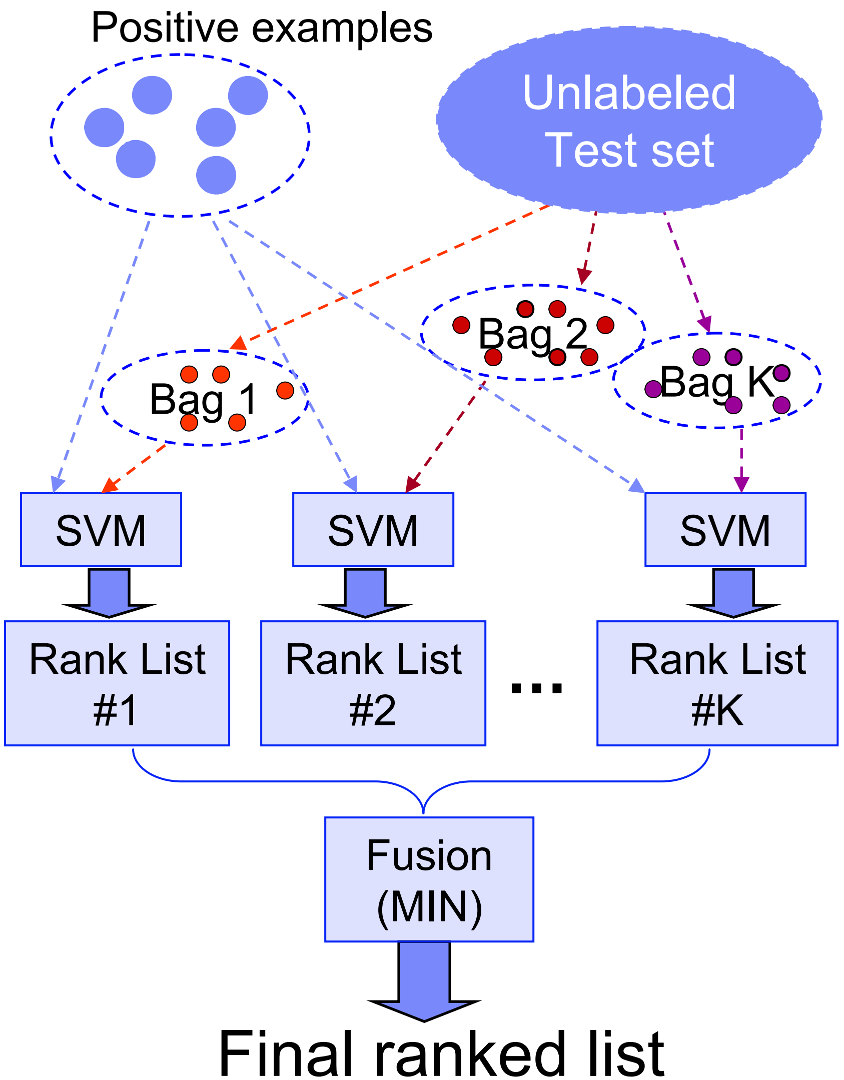
Multi-Example Content Based Retrieval (MECBR)

- **Main idea**
  - Create generalized Boolean CBR from multiple positive query examples
  - Divide topic into multiple “simple” CBR queries and combine results to answer “complex” query
  - Based on modified nearest-neighbor model
- **Implementation**
  - Categorize examples into distinct visual subsets
  - Select representative(s) for each category
  - Execute content-based query with each representative
  - Fuse results within/across categories
- **Parameters**
  - Categorization: clustering, anchoring
  - Representatives: centroid, weighted sampling
  - CBR: features, granularity, normalization
  - Fusion:
    - AND logic within categories (Weighted AVG)
    - OR logic between categories (MAX)



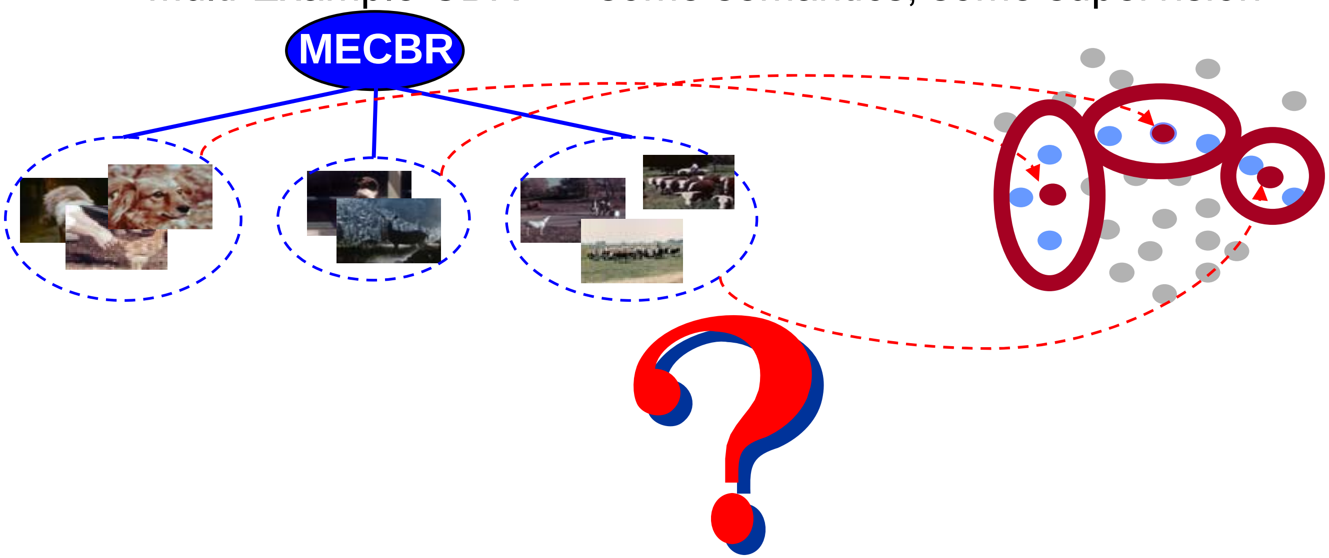
Discriminative Model-Based Retrieval using SVM

- **Main idea**
  - Formulate query as discriminative modeling problem between positive and negative examples
  - Sample pseudo-negatives from unlabeled set
  - Divide problem into multiple sub-problems with simpler decision boundaries—*bagging* approach
  - Intersect “simple” hyperplanes to form “complex” decision boundary for query topic
- **Implementation**
  - Sample N bags of K pseudo-negative examples
  - Build N light-weight SVM models, each separating one bag of negative examples from all positives
  - Rank candidates by their distance to separating hyperplane of each SVM model
  - Fuse all ranked lists (AND fusion) restricting search range to intersection of separating hyperplanes
- **Parameters**
  - Sampling: random, biased, cluster-based, anchoring
  - Bagging: number, size of pseudo-negative bags
  - SVM: kernel parameters

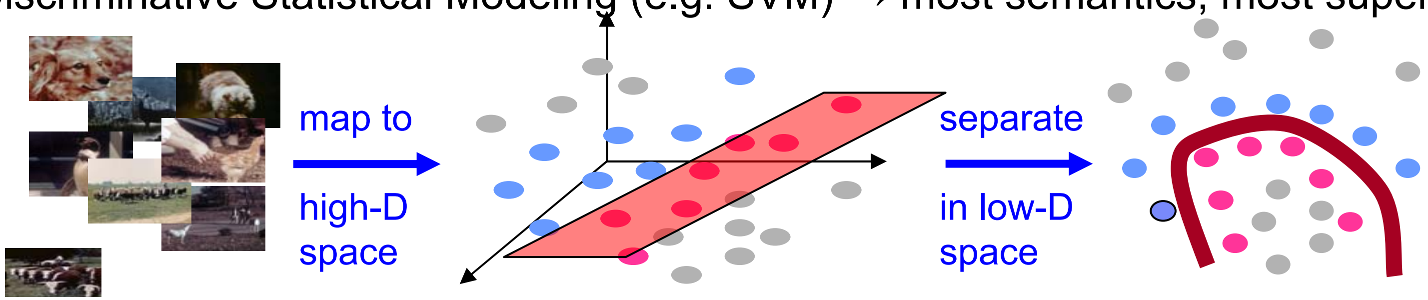


Visual Retrieval Approach—Combination Hypothesis

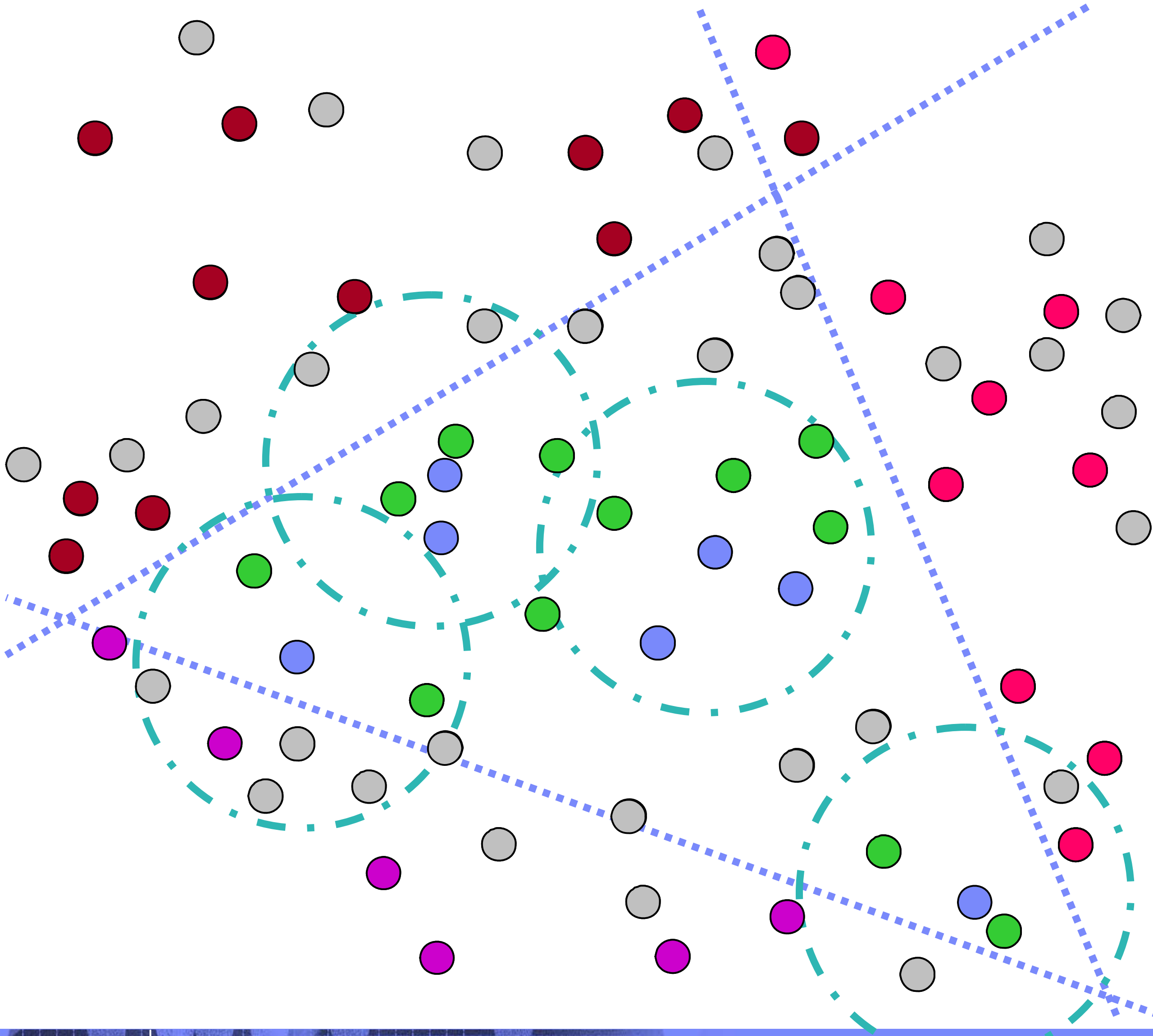
Multi-Example CBR → some semantics, some supervision



Discriminative Statistical Modeling (e.g. SVM) → most semantics, most supervision



Visual Retrieval Approach—Illustration



SVM

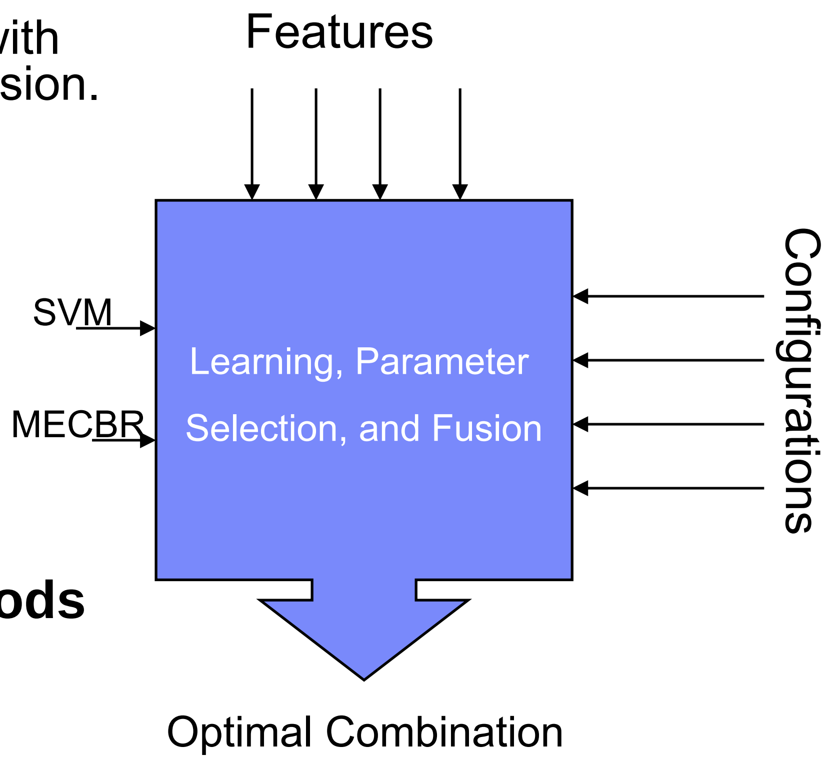
- Selection of pseudo-negative examples based on clustering and sampling
- MIN (AND) fusion of individual rank lists

MECBR

- Selection of good positive examples based on clustering and sampling
- MAX (OR) fusion of individual rank lists

Visual Retrieval Approach—Implementation

- **Combination Hypothesis**
  - We combine two synergistic approaches with different features, and different levels of fusion.
- **Features and granularity**
  - Global Color (Color Correlogram)
  - Spatial Color (Color Moments Grid)
  - Global Texture (Co-occurrence Texture)
  - Spatial Texture (Wavelet Texture Grid)
  - Semantic Features (Model Vectors)
- **Score normalization and fusion methods**
  - Normalization: NONE, RANGE, **STAT**
  - Fusion: MIN, MAX, **AVG**, WAVG
- **Parameter Selection**
  - Query-independent globally-tuned parameters
  - Simple averaging of statistically normalized ranked lists



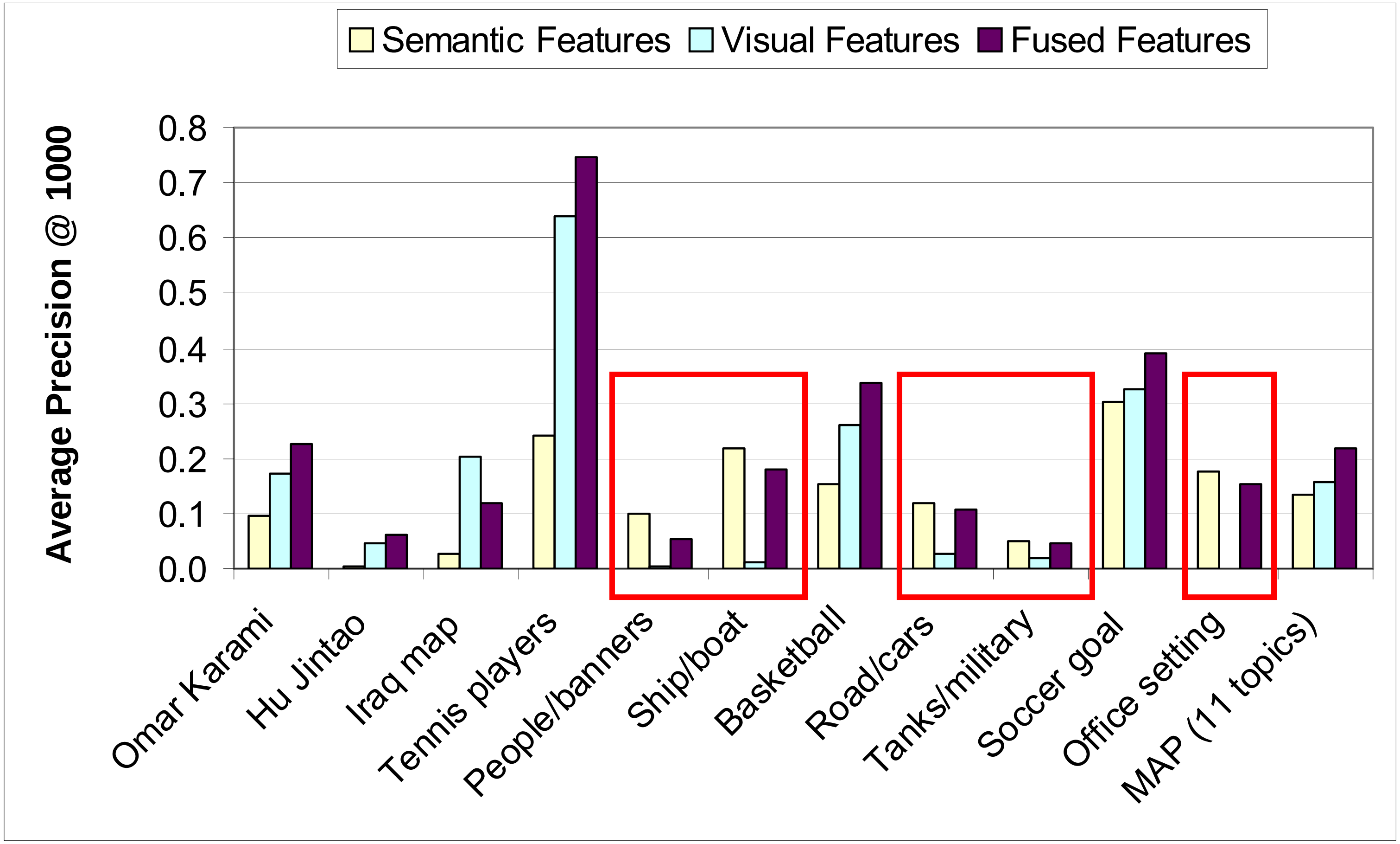
Visual Retrieval Evaluation on TRECVID 2005 Corpus

Feature	MECBR	SVM	Proposed Fusion (% improvement)
Texture	**0.0268**	0.0245	0.0773 (85%)
Global Color	0.0260	**0.0315**	
Spatial Color	0.0210	**0.0418**	
Semantic Feature	**0.0680**	0.0576	0.0693 (2%)
Proposed Fusion (% improvement)	0.1010 (49%)	0.0969 (68%)	0.1101 (42%)

- Mean Average Precision (MAP) computed over the 24 TRECVID 2005 topics
- Parameter-free fusion across features and approaches generalizes extremely well!
- Significant gains in almost all cases!

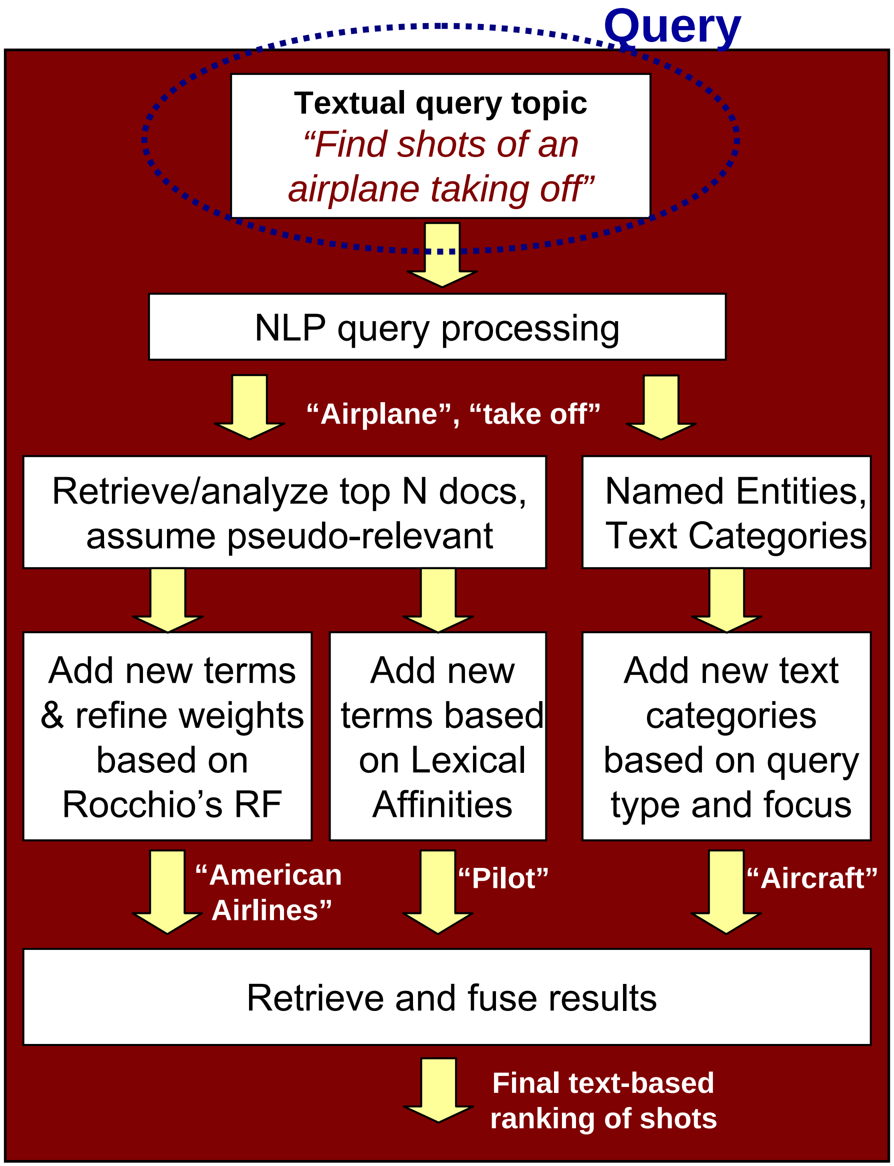
- Improvement over single feature/approach baselines ranges from 62% to over 500%!

Per-Topic Analysis of TRECVID 2005 Visual Retrieval Results



- Visual features work best on sporting events and some named entities
- Semantic features work best on unnamed people and objects

2 IBM Research Text Retrieval System



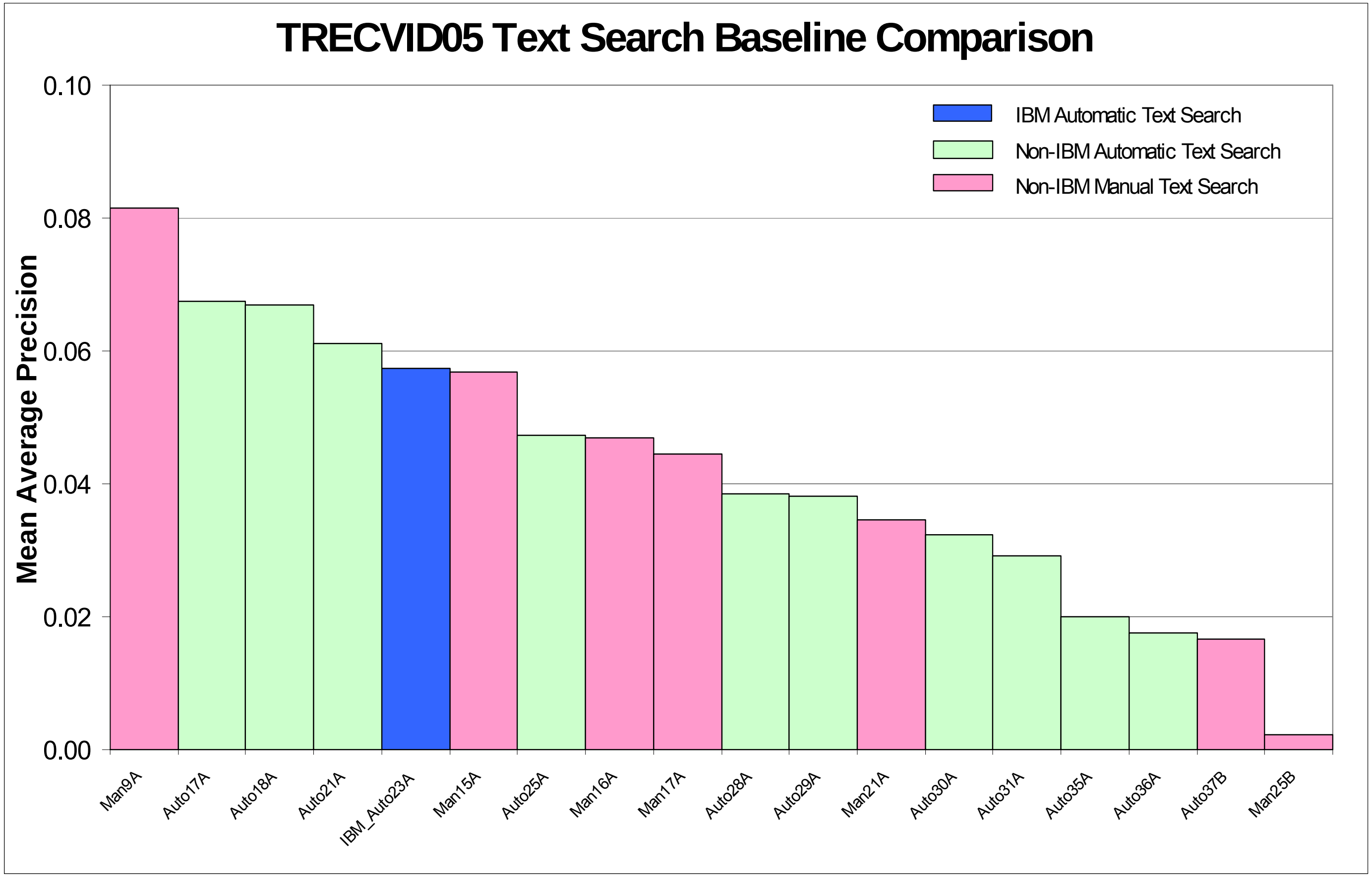
Approaches:

- NLP for query processing:
  - Tokens, stems, phrases
  - POS tagging, query term filtering
  - Named entities & text categories
- Automatic query refinement:
  - Rocchio-based pseudo-RF
  - Lexical affinity-based pseudo-RF
  - QA-based expansion to categories
- Leveraging IBM UIMA SDK

Highlights:

- Competitive to other baselines
- ~40% higher than mean and median
- Better than 6 of 7 manual baselines

TRECVID 2005 Text Retrieval Baseline Performance



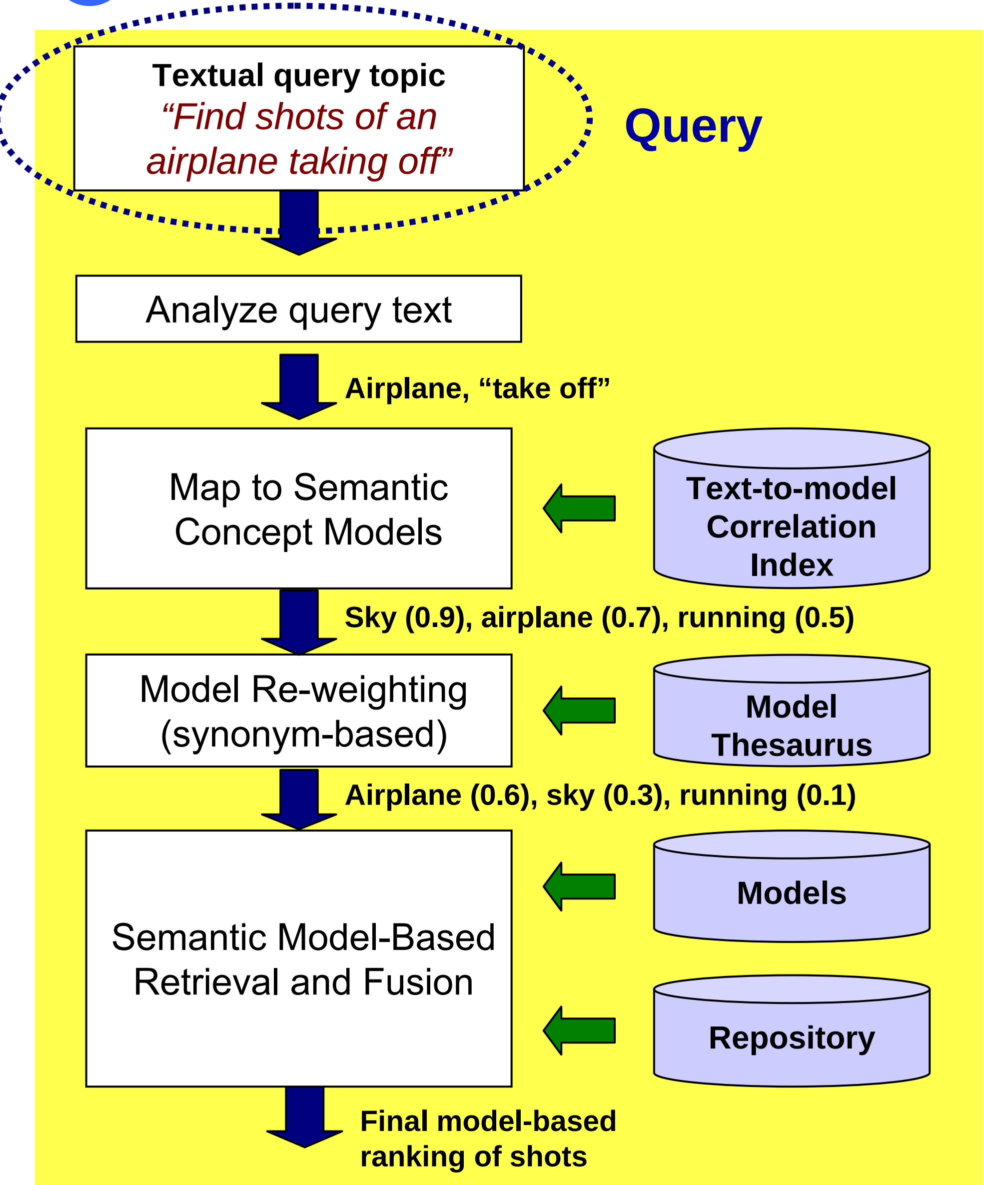
Text Retrieval Results:

IBM automatic:	**0.057**
Best automatic:	**0.067**
Best manual:	**0.081**
Mean and median:	**0.041**

- Automatic text-based retrieval is on par with manual text-based retrieval

3

IBM Research Model-Based Retrieval System



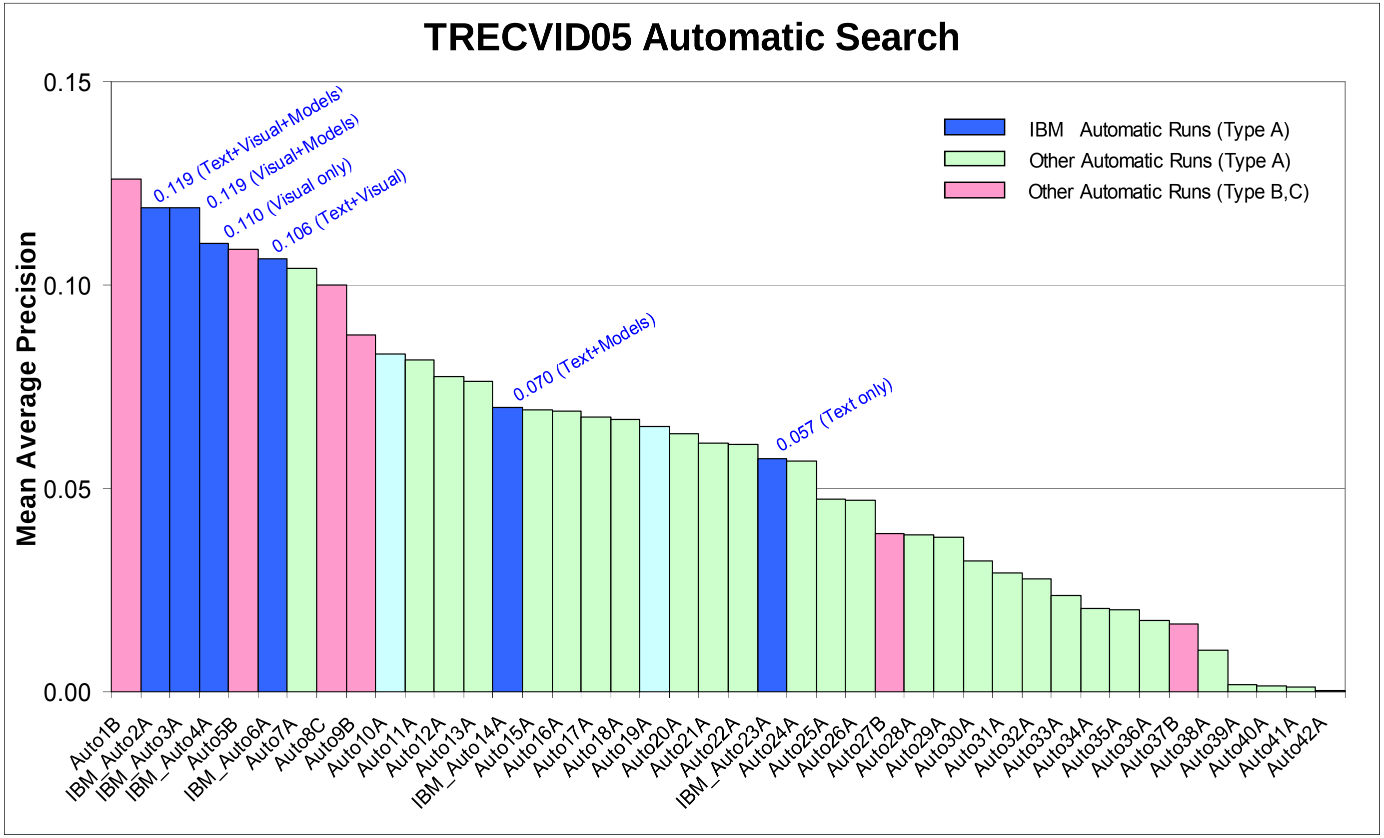
Approaches:

- Automatic mapping of query text to concept models & weights
- Data-driven statistical approach:
  - Co-occurrence statistics between ASR tokens and detected concepts
  - **Supervised**: learn correlations using concept ground truth (training set)
  - **Unsupervised**: learn correlations using concept detection confidences (test set)
- Language-driven lexical approach

Highlights

- Used to re-rank text/visual baselines
- Improved baselines by 20-60%

Comparison of Automatic Search Approaches



Submitted Runs:

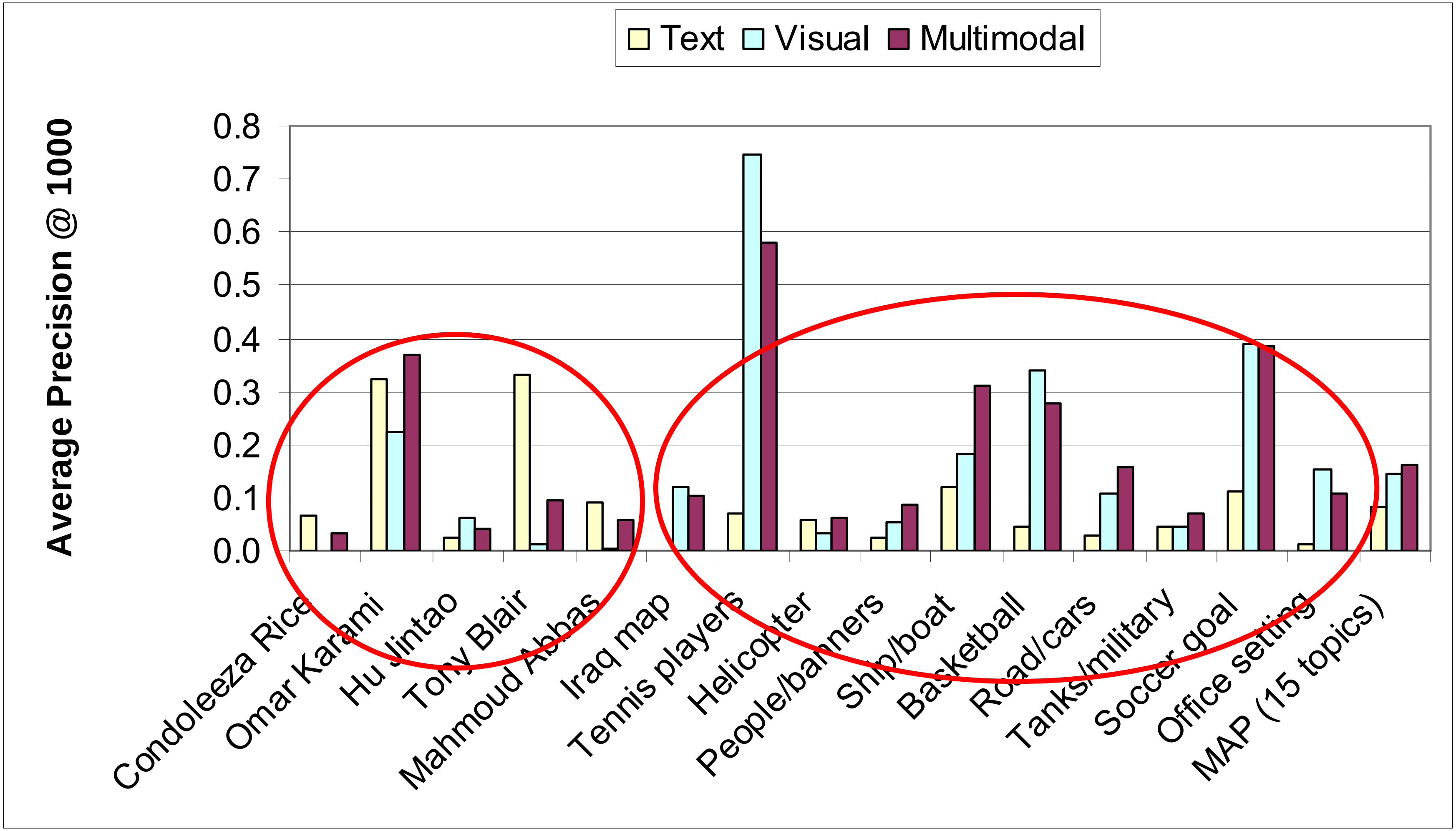
Text:	**0.057**
Text+Models:	**0.070**
Visual:	**0.110**
Visual+Models:	**0.119**
Multimodal:	**0.106**
Multimodal+Models:	**0.119**

Not Submitted Runs:

Text:	**0.060**
Text+Models:	**0.100**
Multimodal:	**0.101**
Multimodal+Models:	**0.134**
Oracle selection**:**	**0.180**

- Visual modality outperforms speech modality by a large margin (2x)
- Model-based re-ranking improves on both modalities (20-60%) with 39 models only
- Outperform other automatic type A search approaches and all but 2 of 26 manual runs!

Per-Topic Analysis of IBM Automatic Search Methods



Summary

- Automatic query expansion for text retrieval
  - Competitive text-only baseline but much worse compared to visual modality
- Light-weight learning for visual retrieval
  - Successfully applied SVMs to search scenarios with very few examples
  - Combination hypothesis across complementary approaches and features
    - Worked extremely well even with simple query-independent fusion
  - Best overall approach for automatic type A search this year
- Statistical and lexical model-based retrieval and re-ranking
  - Fully automatic model selection and weighting from query text or examples
  - Dramatically improved text and visual baselines (20-60% gain with 39 models)

Observations

- Conclusions
  - Visual modality 2x better than speech modality this year
  - Concept models helped significantly (>50% gain over baselines)
  - Automatic search on par with manual search!

- Proposals
  - Use recall (or precision) at certain depth for evaluation of automatic search (approximates interactive search)
  - Change guidelines for manual search (merge w/ automatic?)