TRECVID-2004: Shot Boundary Detection Task Overview

Alan Smeaton
Dublin City University
&
Paul Over, Joaquim Arlandis
NIST

SB Task Definition

- Shot boundary detection is a fundamental task in any kind of video content manipulation;
- Goal: To further progress the task of SB detection in video;
- Secondary goal is to provide a stepping-stone for some groups who wish to “break into” video retrieval;
- Task is to identify the shot boundaries with their location and type (cut or gradual) in the given video clip(s)

SB Task Details

- Groups may submit up to 10 runs + 10 additional runs which ALL use as additional system input the LIMSI ASR (RMIT did this);
- Comparison to human-annotated reference (thanks to Jonathan Lasko, again);
- Groups were asked to provide some standard information on the processing complexity of each run;

2004: Shot boundary task groups (17, from 12)

AIIA Laboratory	Greece	SB	--	--	--
CLIPS-LSR-LIS	France	SB	SS	FE	SE
Fraunhofer (Heinrich Hertz) Institute	Germany	SB	--	--	--
FX Palo Alto Laboratory	US	SB	--	--	SE
IBM Research	US	SB	SS	FE	SE
Imperial College, London	UK	SB	SS	FE	SE
KDDI R&D Laboratories	Japan	SB	SS	--	--
RMIT University	Australia	SB	SS	--	--
SAMOVA/IRIT/UPS	France	SB	--	--	--
Tsinghua National Laboratory for I & T	China	SB	--	FE	--
University of Bremen/TZI	Germany	SB	--	--	--
University of Bordeaux	France	SB	--	--	--
University of Iowa	US	SB	SS	FE	--
Queen Mary, University of London	UK	SB	--	--	SE
University of Maryland	US	SB	--	--	--
University of Sao Paolo/IME	Brazil	SB	--	--	--
University Rey Juan Carlos	Spain	SB	--	--	--

2004 Shot boundary data

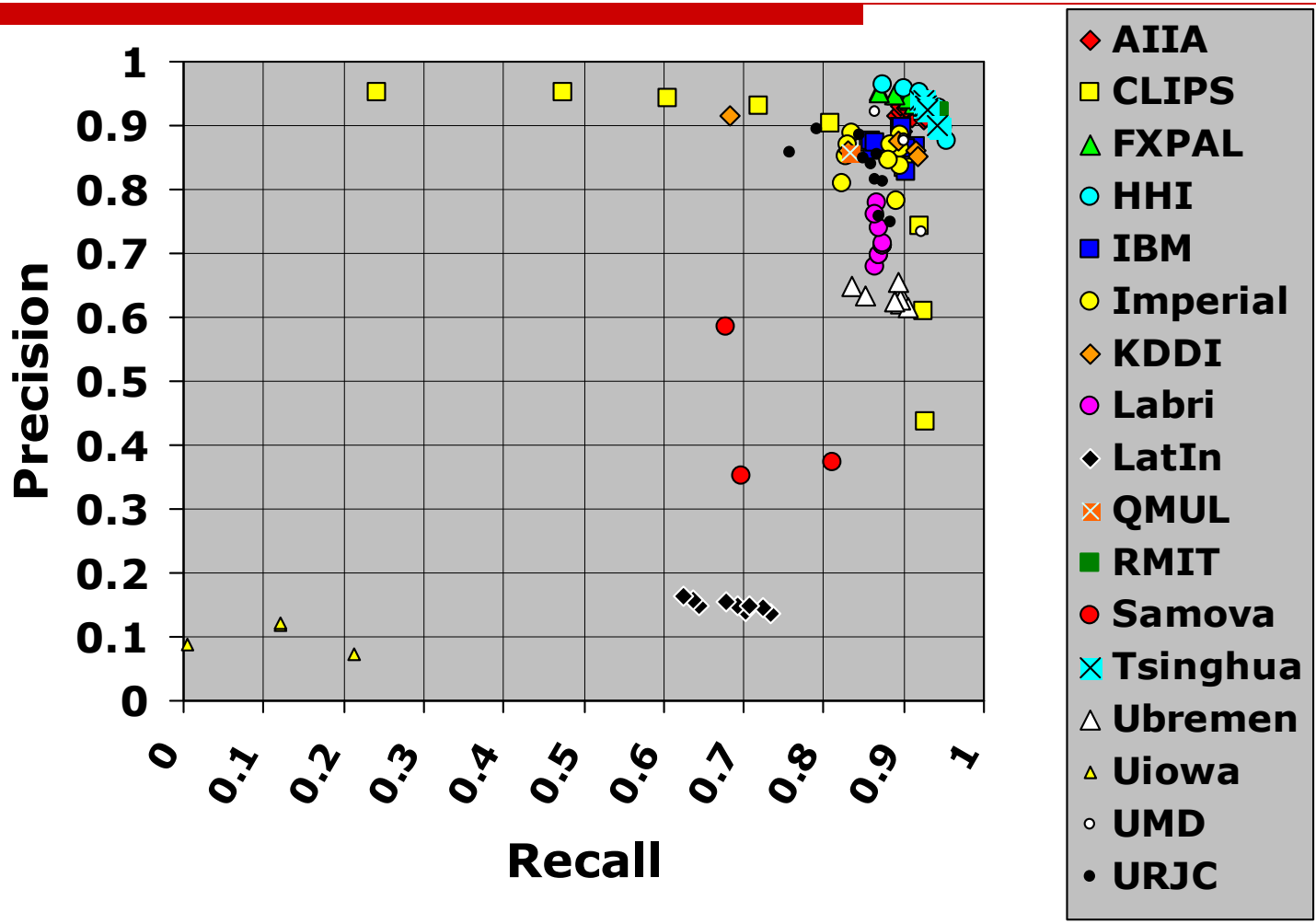
- 12 randomly selected 30-minute videos
- Total frames = 618,409
- Total transitions = 4,806
- 0.777 transitions/100 frames (< than in 2003)
- Transition types:
  - **2,774 (57.7%) Cuts**
  - **1,525 (31.7%) Dissolves (> than in 2003)**
  - **230 (4.8%) Fade-out/-in**
  - **276 (5.7%) other**

2003 Shot boundary data

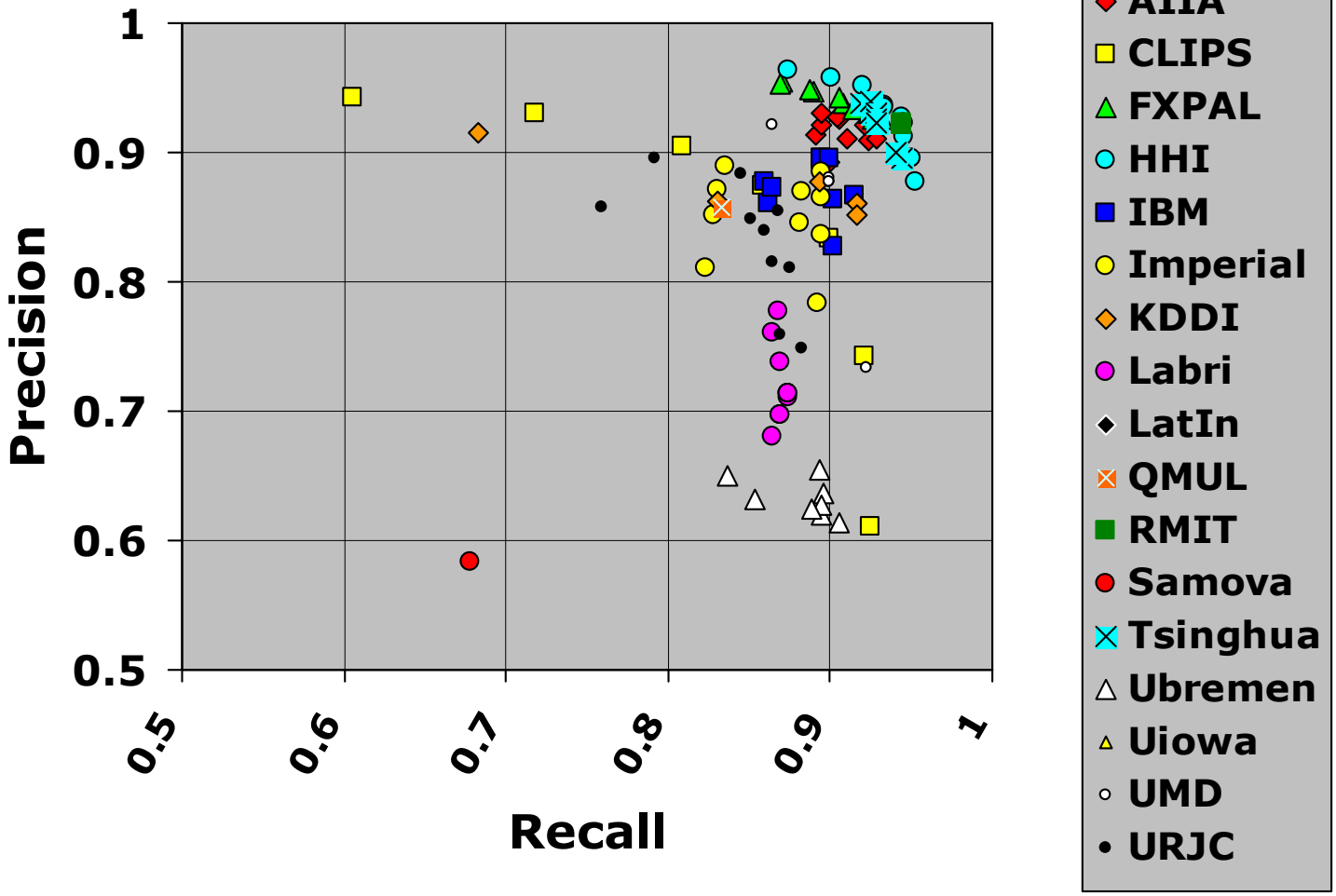
- ABC/CNN + C-SPAN data
- Total frames = 596,054 (416,128 ABC/CNN)
- Total transitions = 3,734 (3,574 ABC/CNN)
- 0.626 transitions/100 frames (0.859 ABC/CNN)
- Transition types:

	ABC/CNN only	
2,644 (70.7%)	2488 (69.6%)	Cuts
753 (20.2%)	749 (21.0%)	Dissolves
116 (3.1%)	116 (3.2%)	Fadeout/in
221 (5.9%)	221 (6.2%)	other

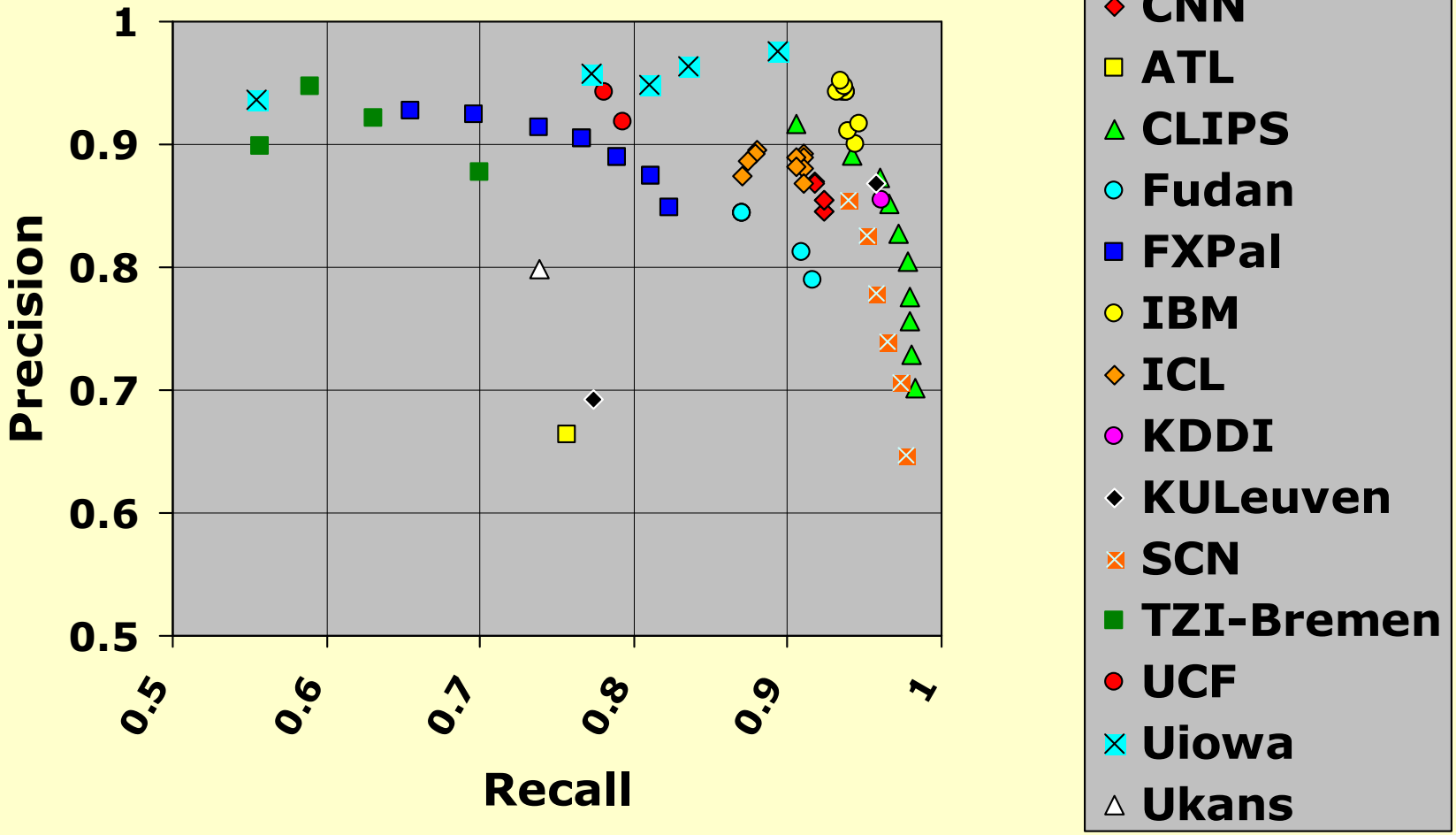
2004: Cuts


Precision
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Recall
AIIA
CLIPS
FXPAL
HHI
IBM
Imperial
KDDI
Labri
LatIn
QMUL
RMIT
Samova
Tsinghua
Ubremen
Uiowa
UMD
URJC

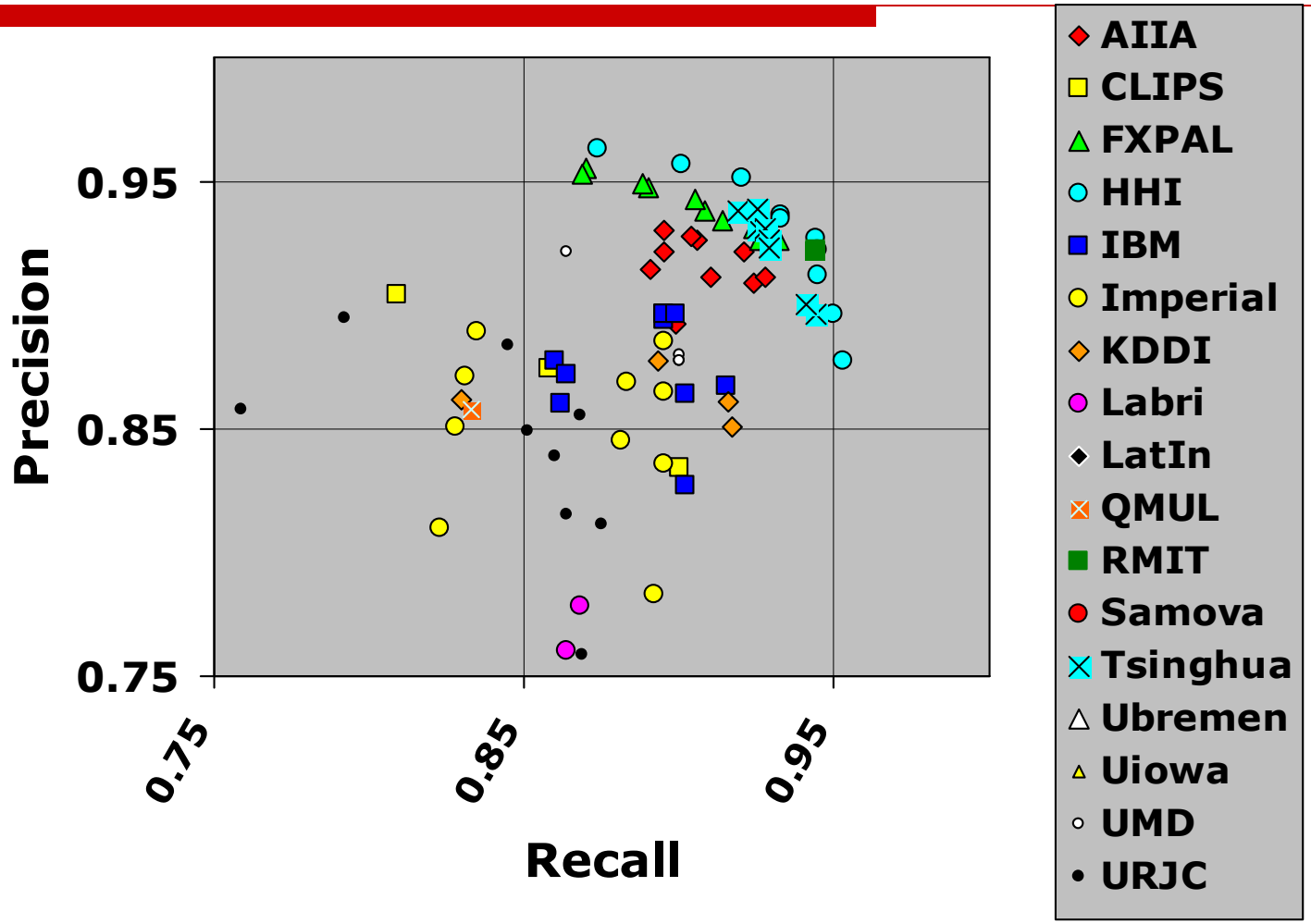
2004: Cuts (zoomed)



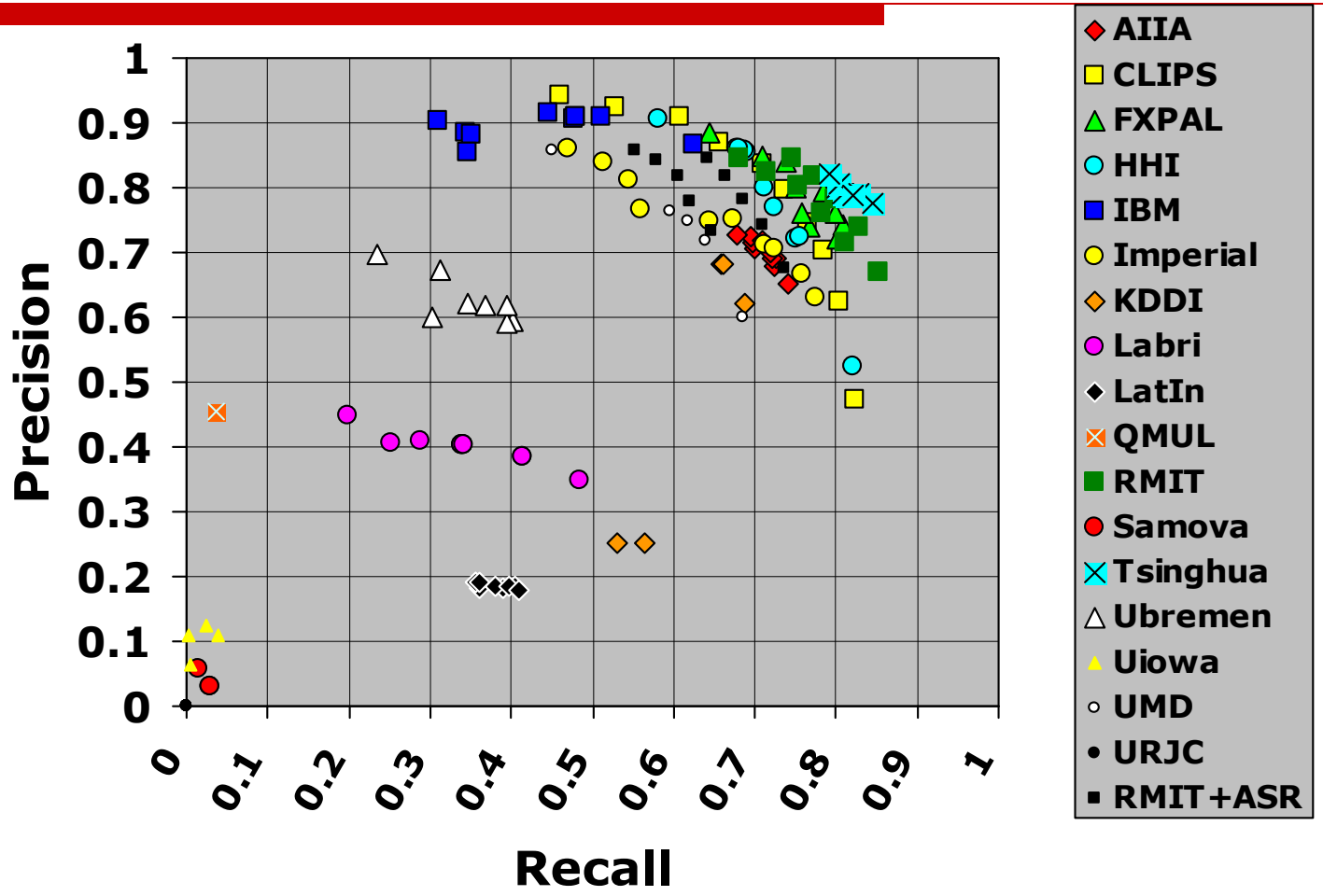
2003: Cuts (zoomed)



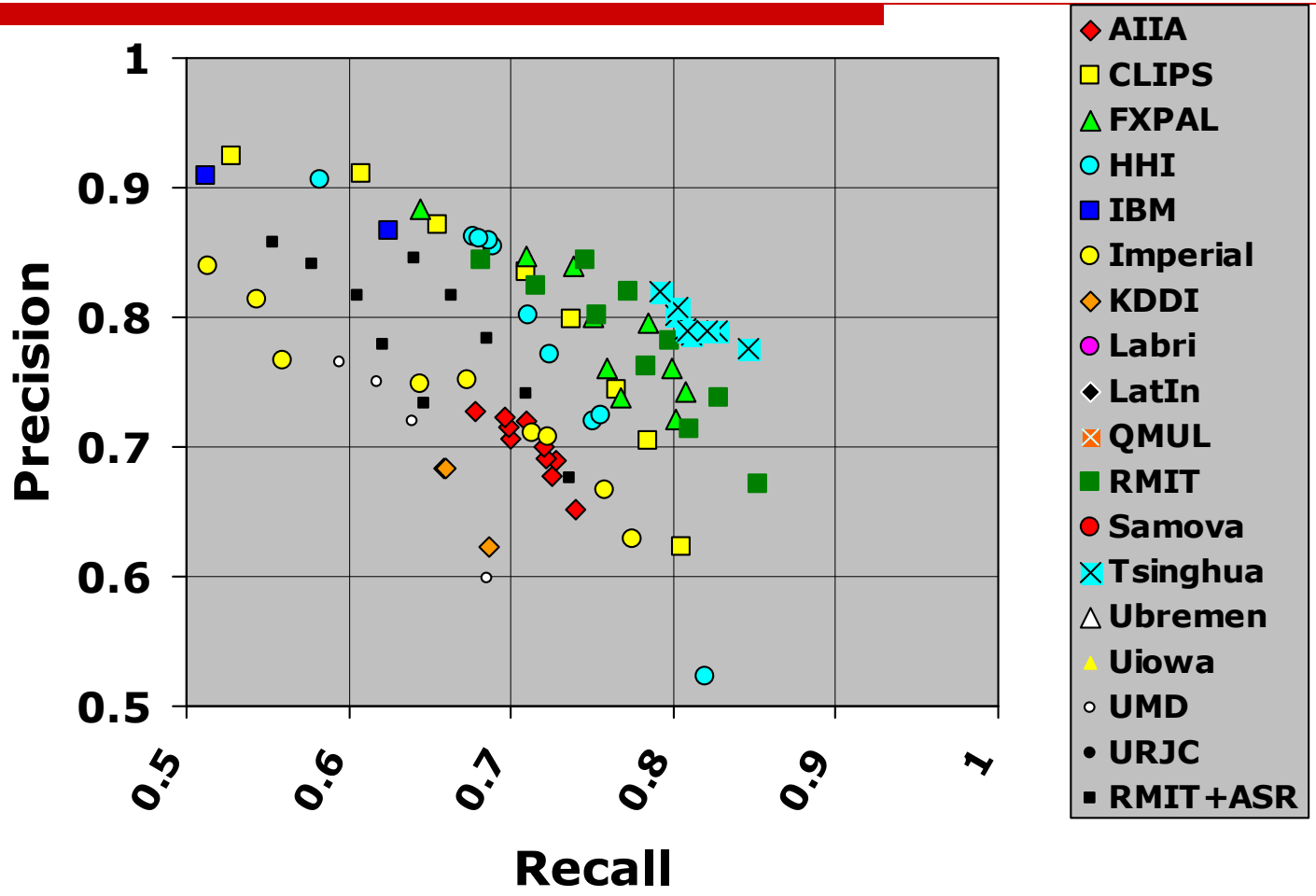
2004: Cuts (zoomed again)



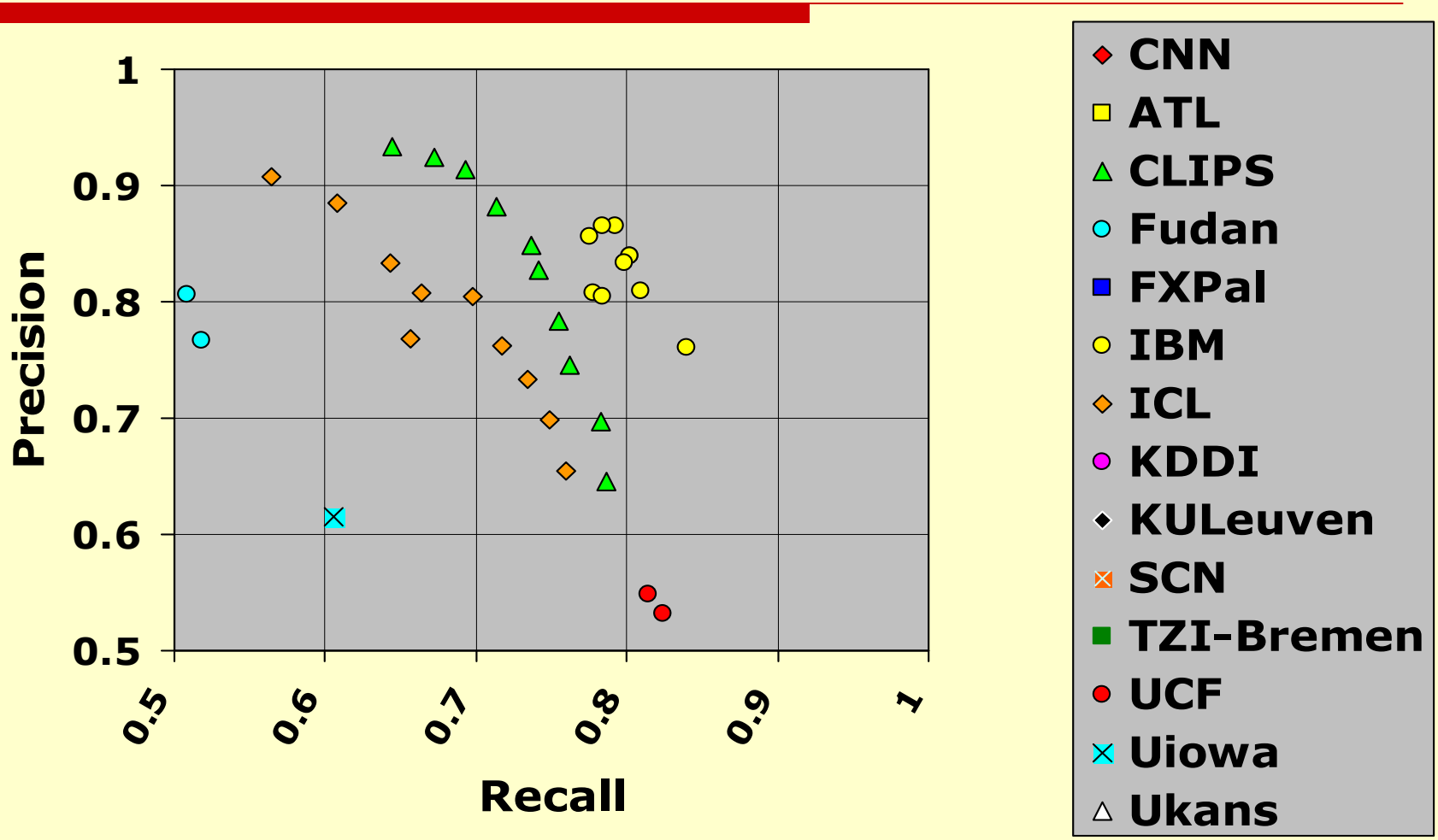
2004: Gradual transitions



2004: Gradual transitions (zoomed)



2003: Gradual Transitions (zoomed)



Evaluation Measures

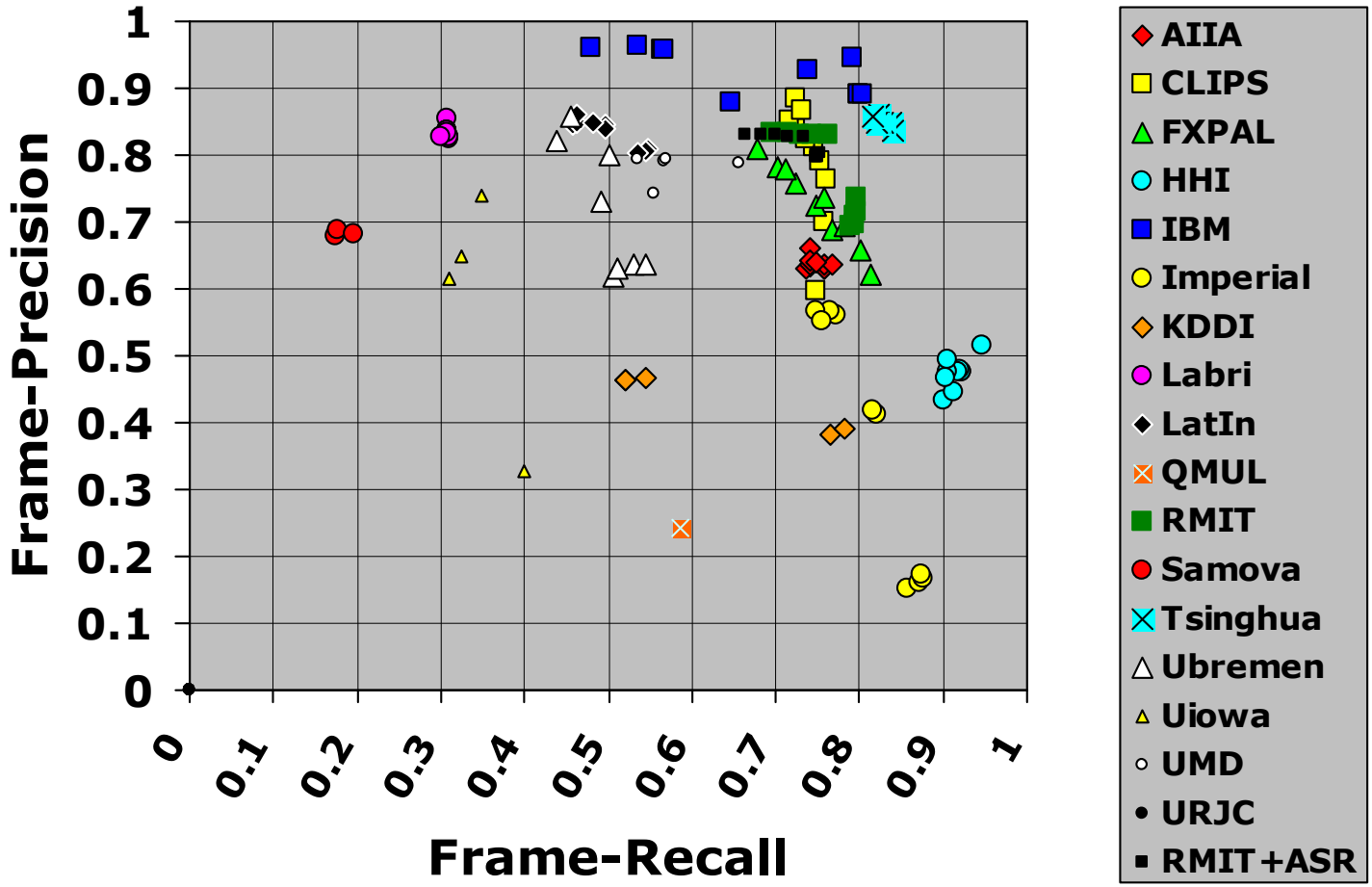
$$\text{Precision} = \frac{\#\ \text{Transitions Correctly Reported}}{\#\ \text{Transitions Reported}}$$

$$\text{Recall} = \frac{\#\ \text{Transitions Correctly Reported}}{\#\ \text{Transitions in Reference}}$$

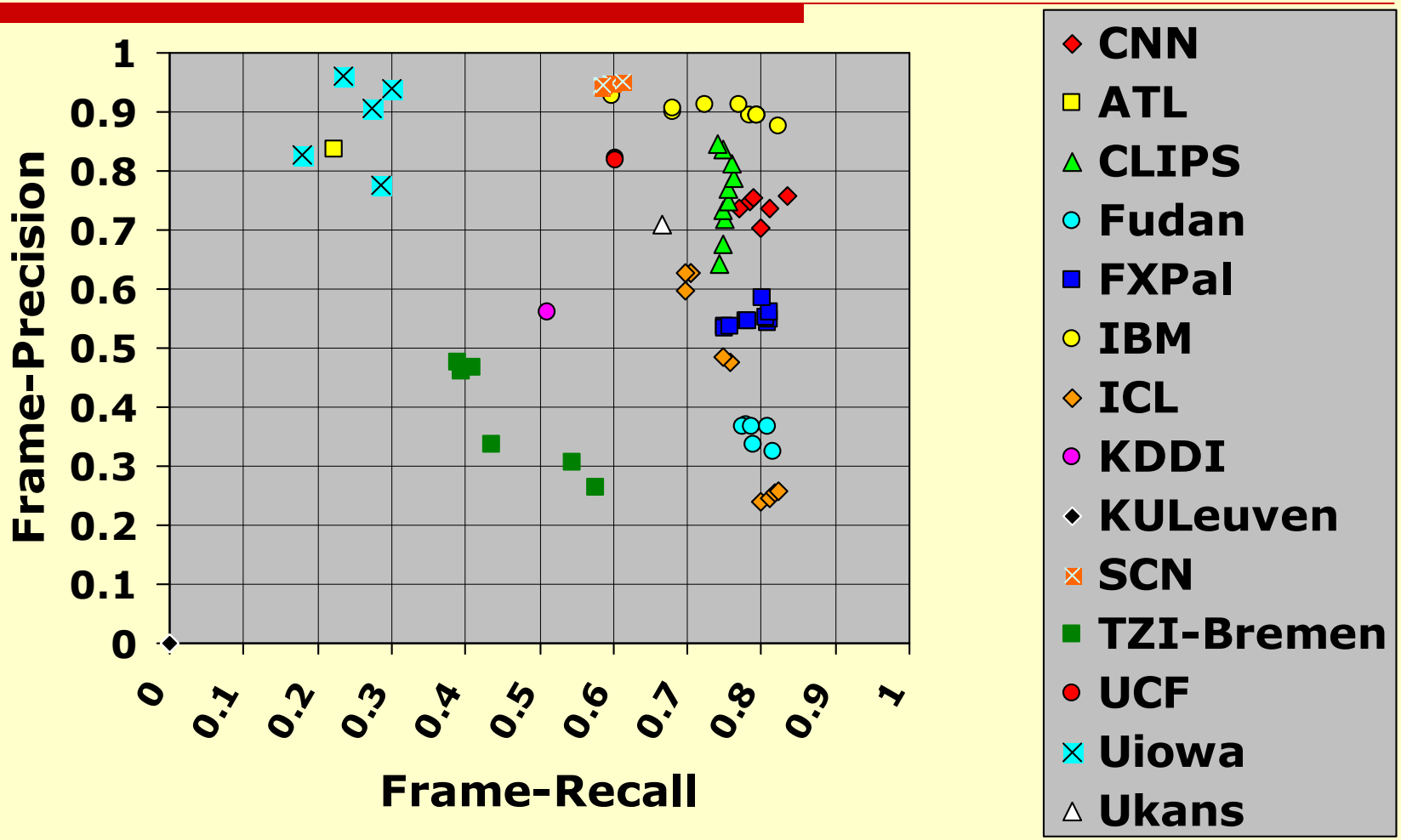
$$\text{Frame Precision} = \frac{\#\ \text{Frames Correctly Reported in Detected Transitions}}{\#\ \text{Frames reported in Detected Transitions}}$$

$$\text{Frame Recall} = \frac{\#\ \text{Frames Correctly Reported in Detected Transitions}}{\#\ \text{Frames in Reference Data for Detected Transitions}}$$

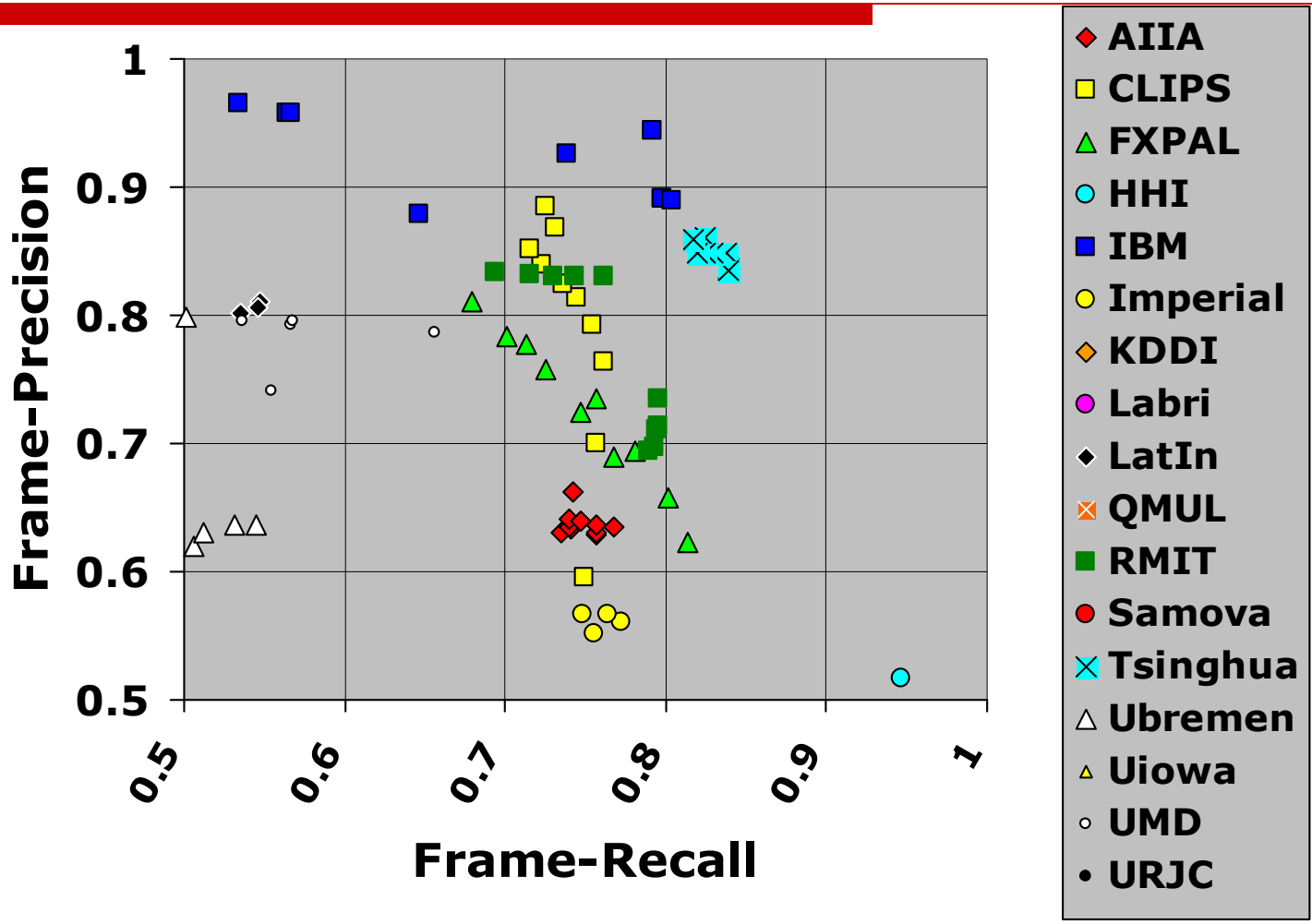
2004: Gradual Transitions (Frame-P&R)



2003: Gradual Transitions (Frame-P&R)



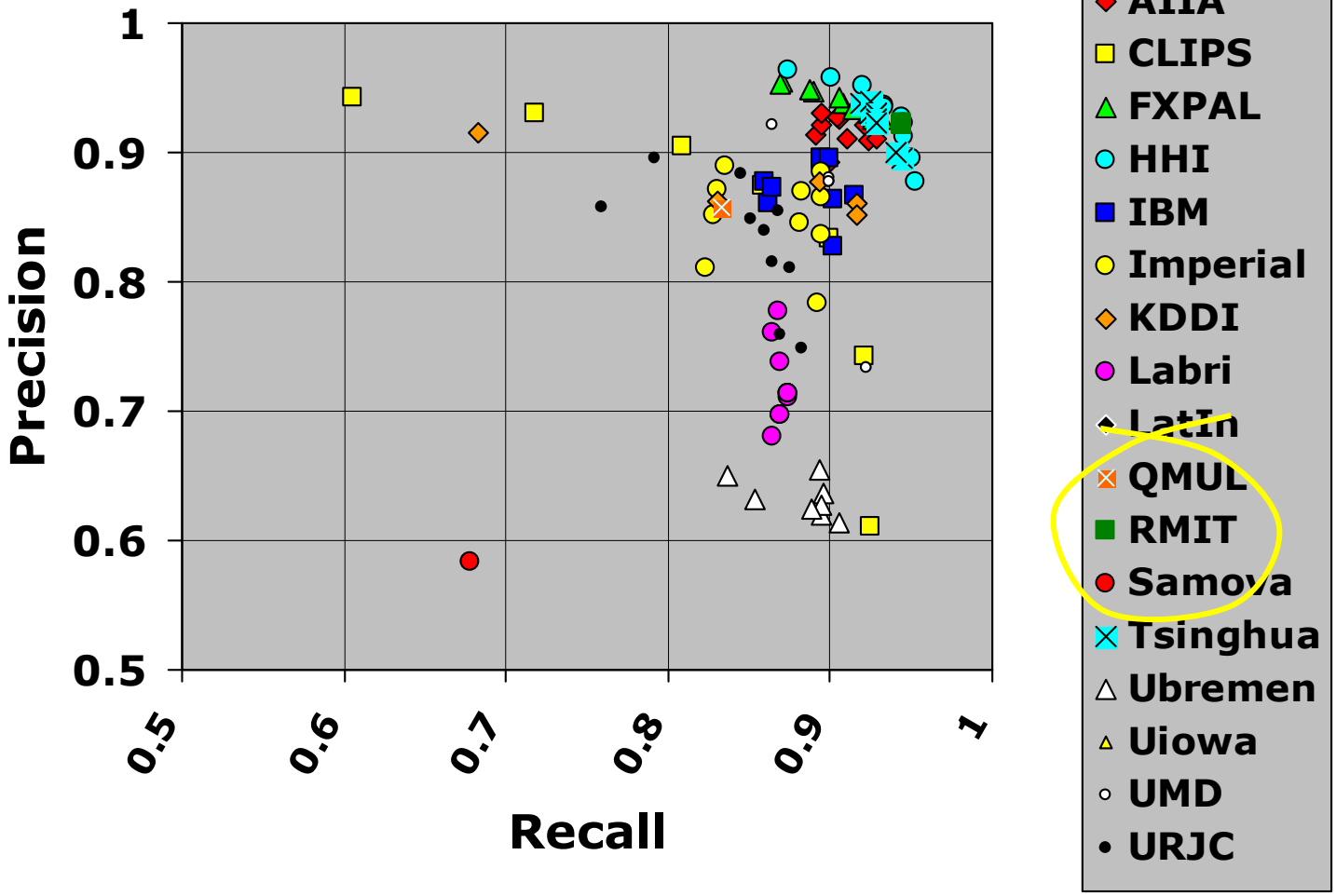
2004: Gradual transitions (zoomed)



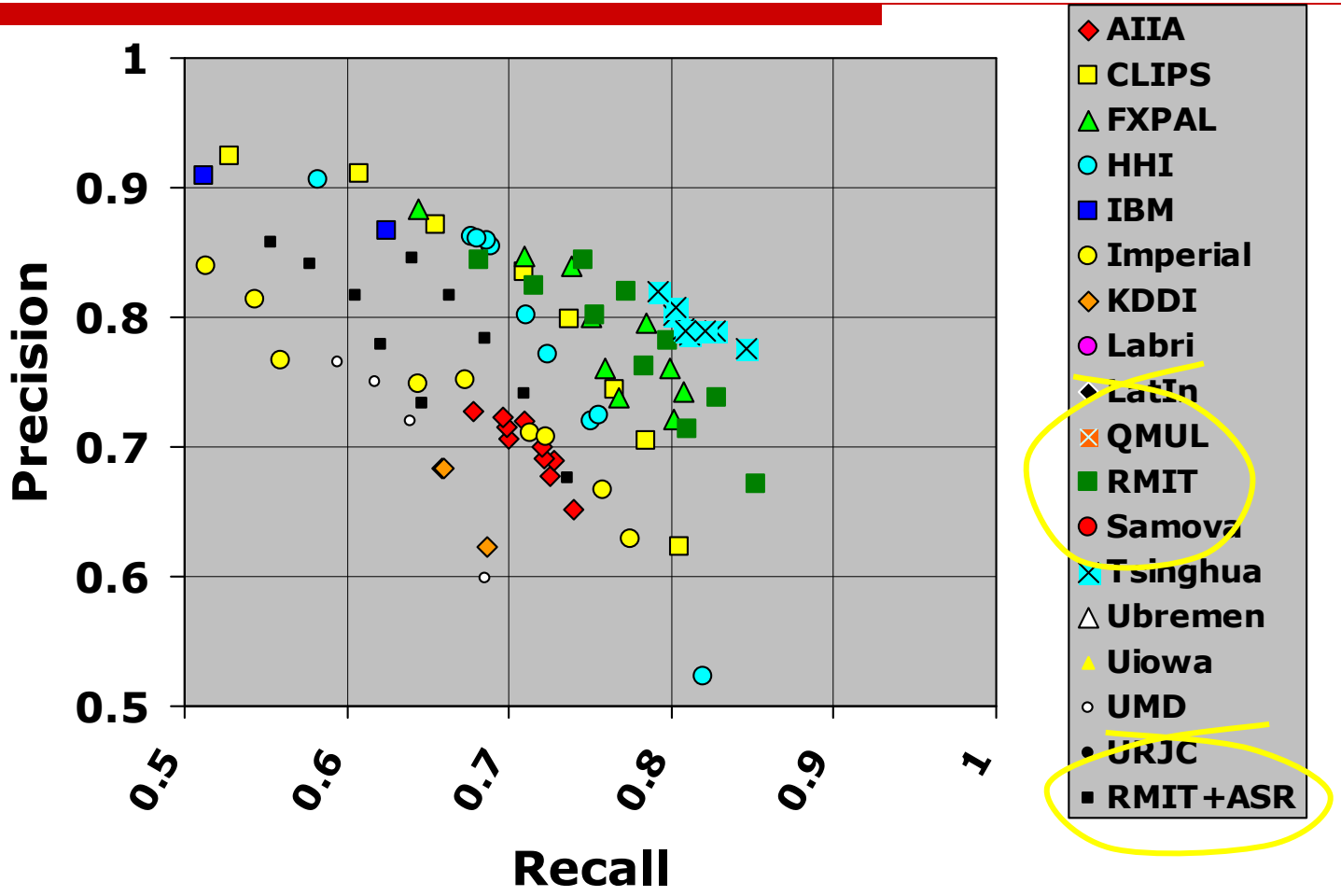
RMIT: Headlines - (to speak following)

- Approach
  - 10 runs + 10 runs with ASR;
  - A new post-processing step to GT detection scheme from TRECVid 2003;
- Features
  - Post-processing improves GT detection;
  - ASR was used in post-processing to remove GTs that coincide with spoken words, but ineffective as it dramatically reduces recall while precision remains;
- Performance
  - Using an Intel P4 / 3.0 GHz, runs took 15.5k seconds, 99% spent on decoding, 1% on segmentation, close to RT;
- Results

2004: Cuts (zoomed)



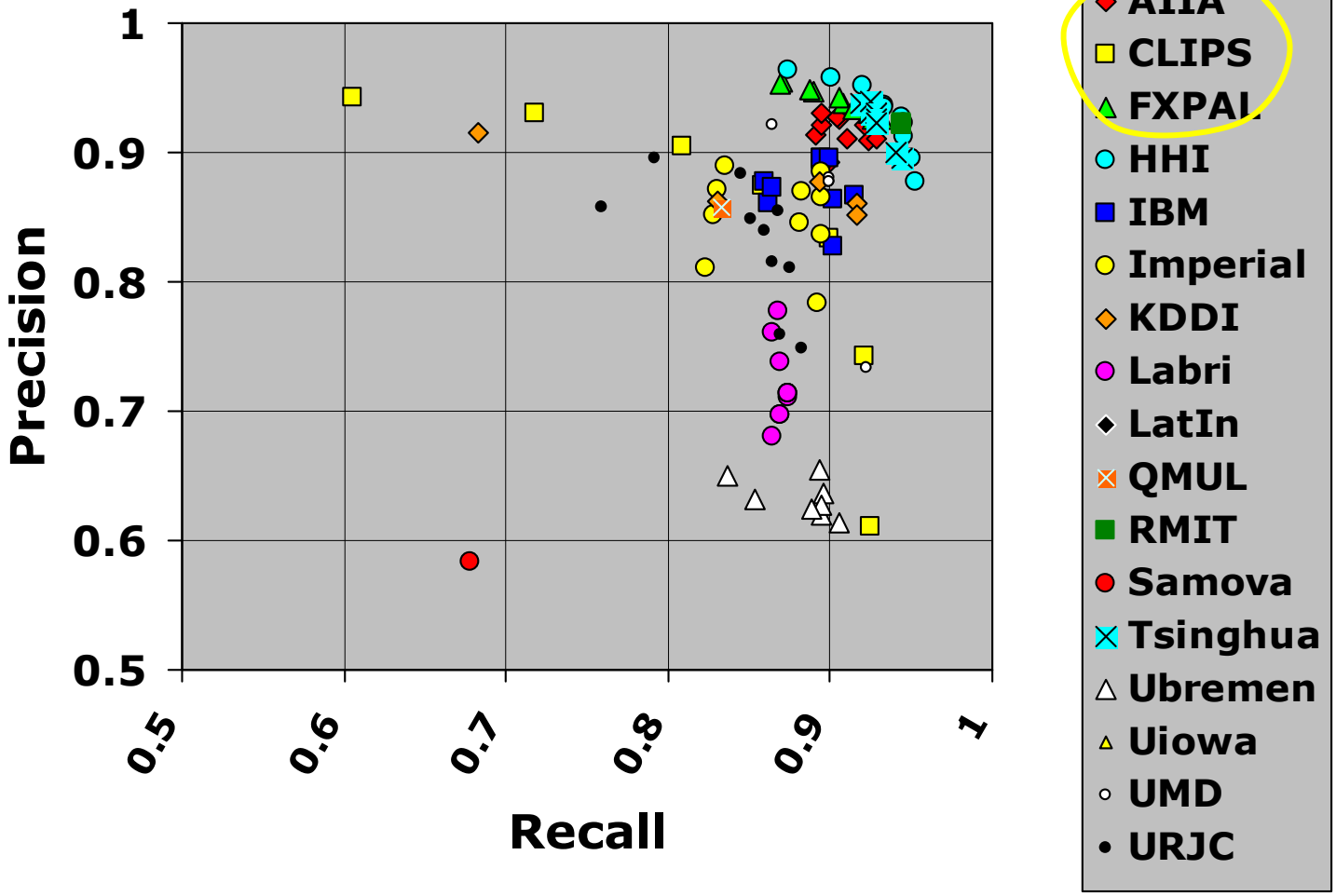
2004: Gradual transitions (zoomed)



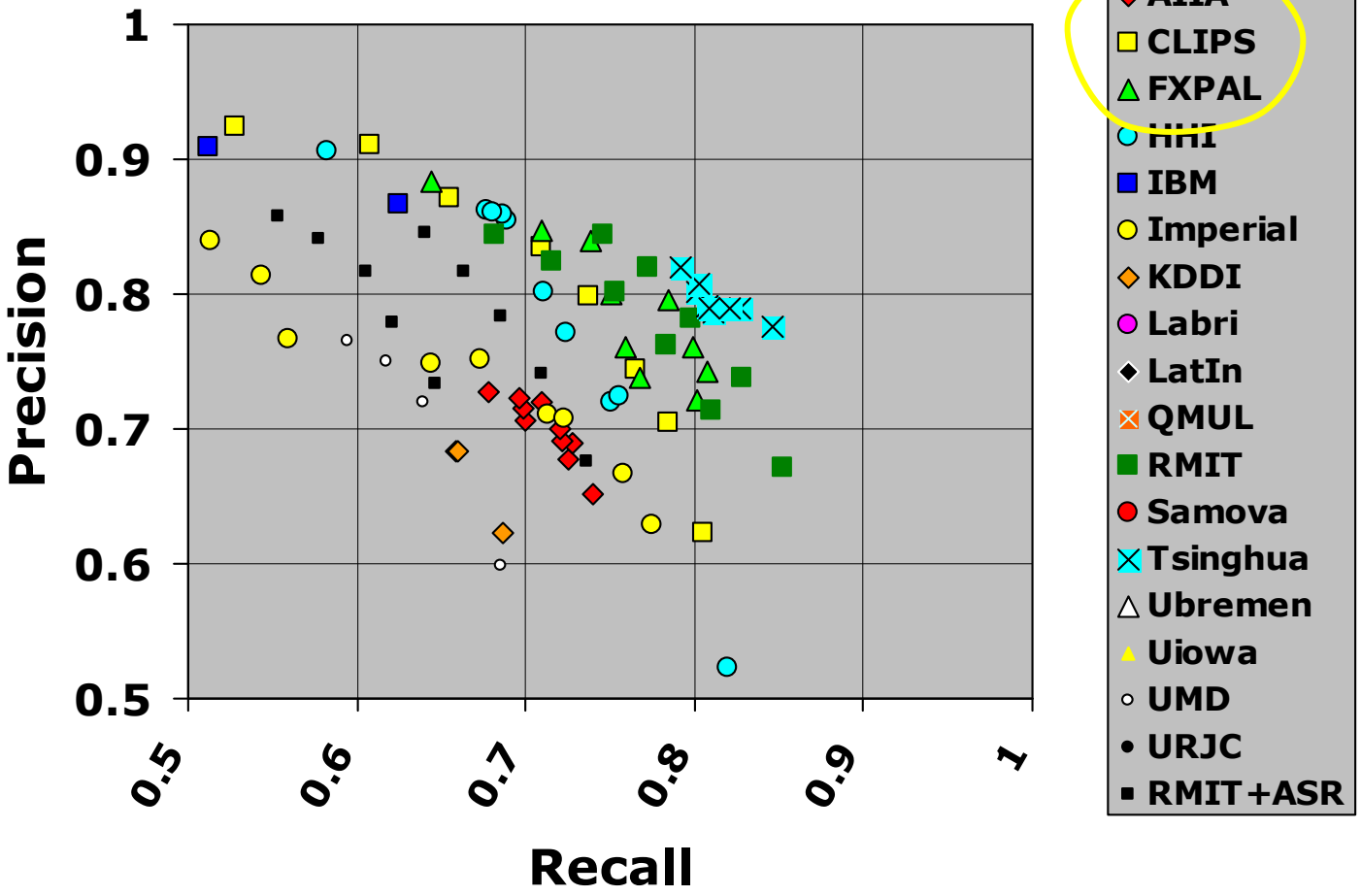
CLIPS: Headlines

- Approach
  - Same as previous years;
- Features
  - Detect cuts by image comparisons after motion compensation and GTs by comparing norms of first and second temporal derivatives of the images;
- Performance
  - Using an Intel P4 / 3.2 GHz, most runs were 15k seconds, faster to decode than RMIT, segmentation was longer, and overall about the same, quicker than RT;
- Results

2004: Cuts (zoomed)



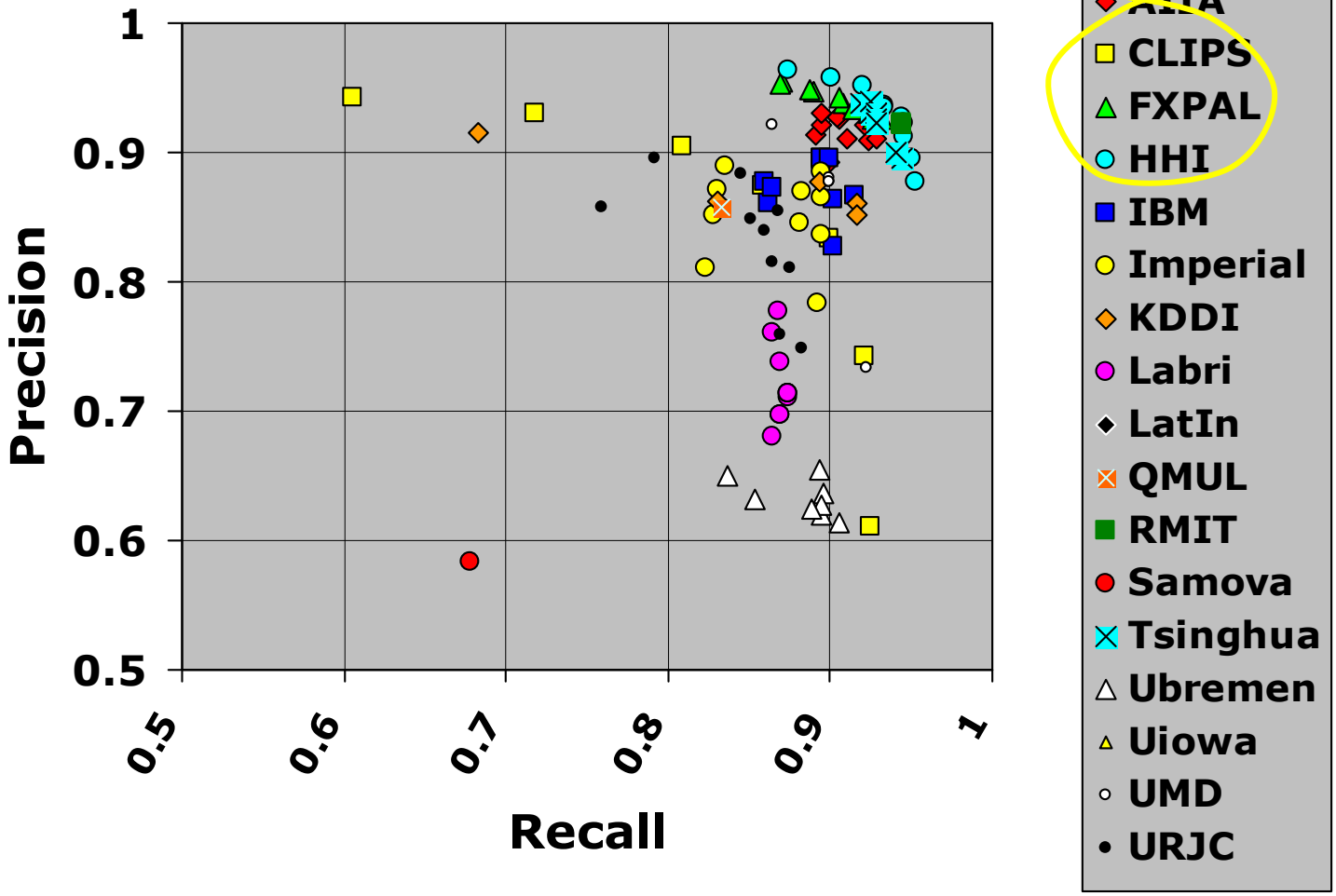
2004: Gradual transitions (zoomed)



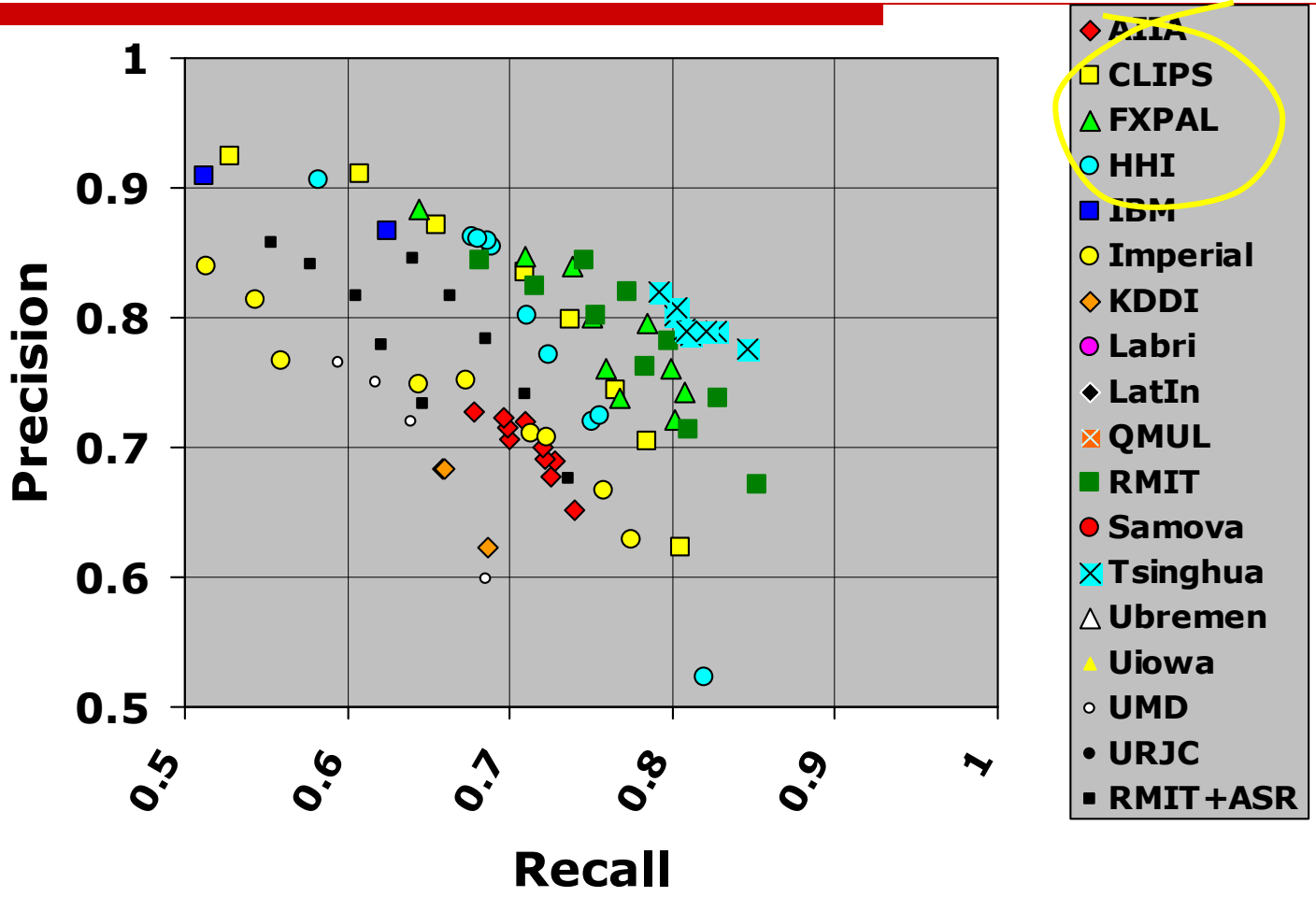
FXPAL: Headlines - (to speak following)

- Approach
  - Combine pairwise similarities between images in the locality and supervised classification;
- Features
  - TRECVid2003 did the multiple pairwise similarities and the classification is new for 2004;
- Performance
  - Using an AMD 64 3500, runs took 45k seconds, expensive, more time spent on decoding than on segmentation;
- Results

2004: Cuts (zoomed)



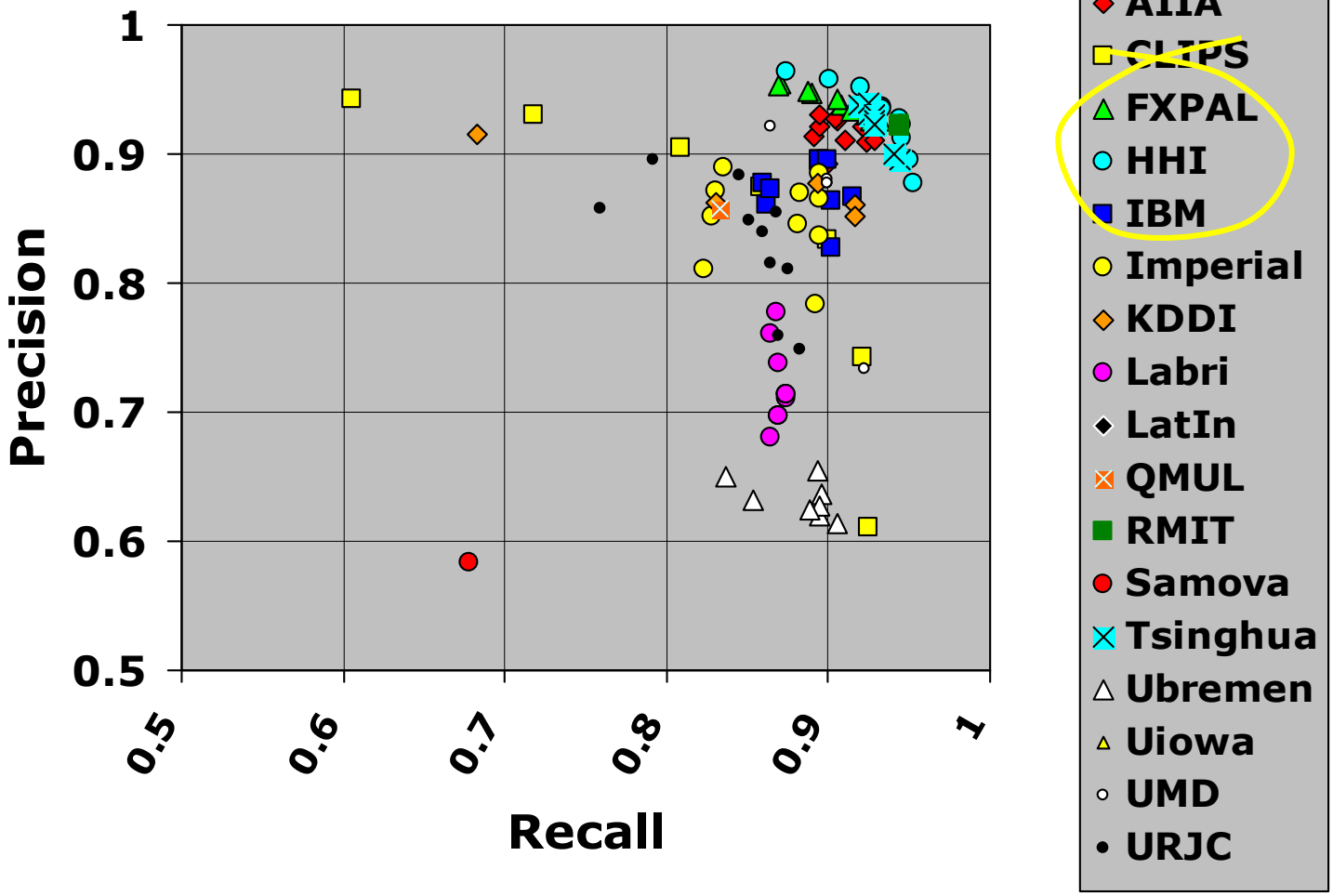
2004: Gradual transitions (zoomed)



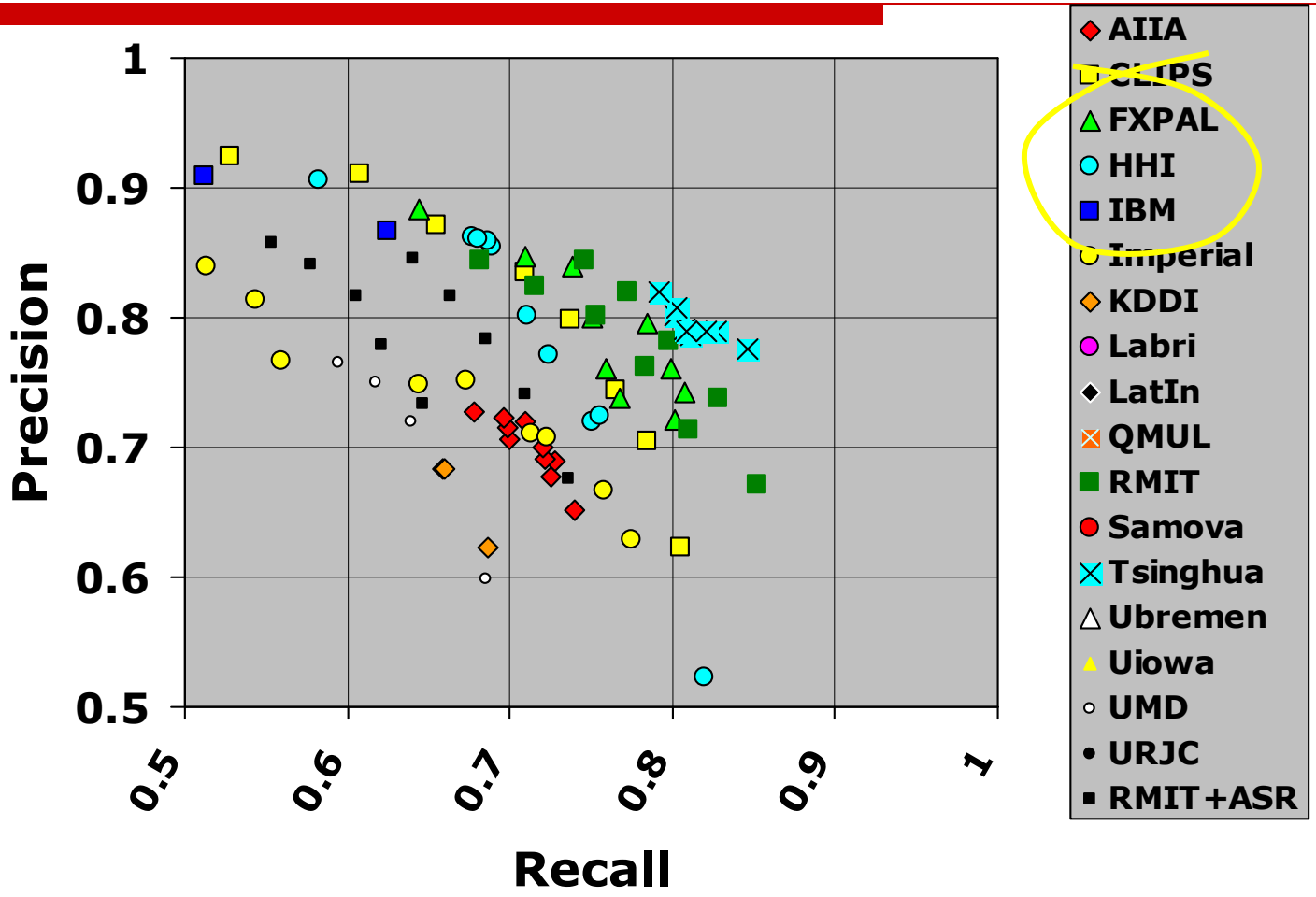
HHI: Headlines

- Approach
  - Cuts: pixel and edge differences with adaptive thresholding, photo and slo-mo detection also;
  - GTs: edge energy statistics, pixel and histogram differences, and wipe detector;
- Features
  - New to TRECVid, excellent performance;
- Performance
  - Using an Intel P4 / 3.06 GHz, runs took c.1k seconds, fast GPL public decoder (60% time) and segmentation took 40% - 20x faster than real time;
- Results

2004: Cuts (zoomed)



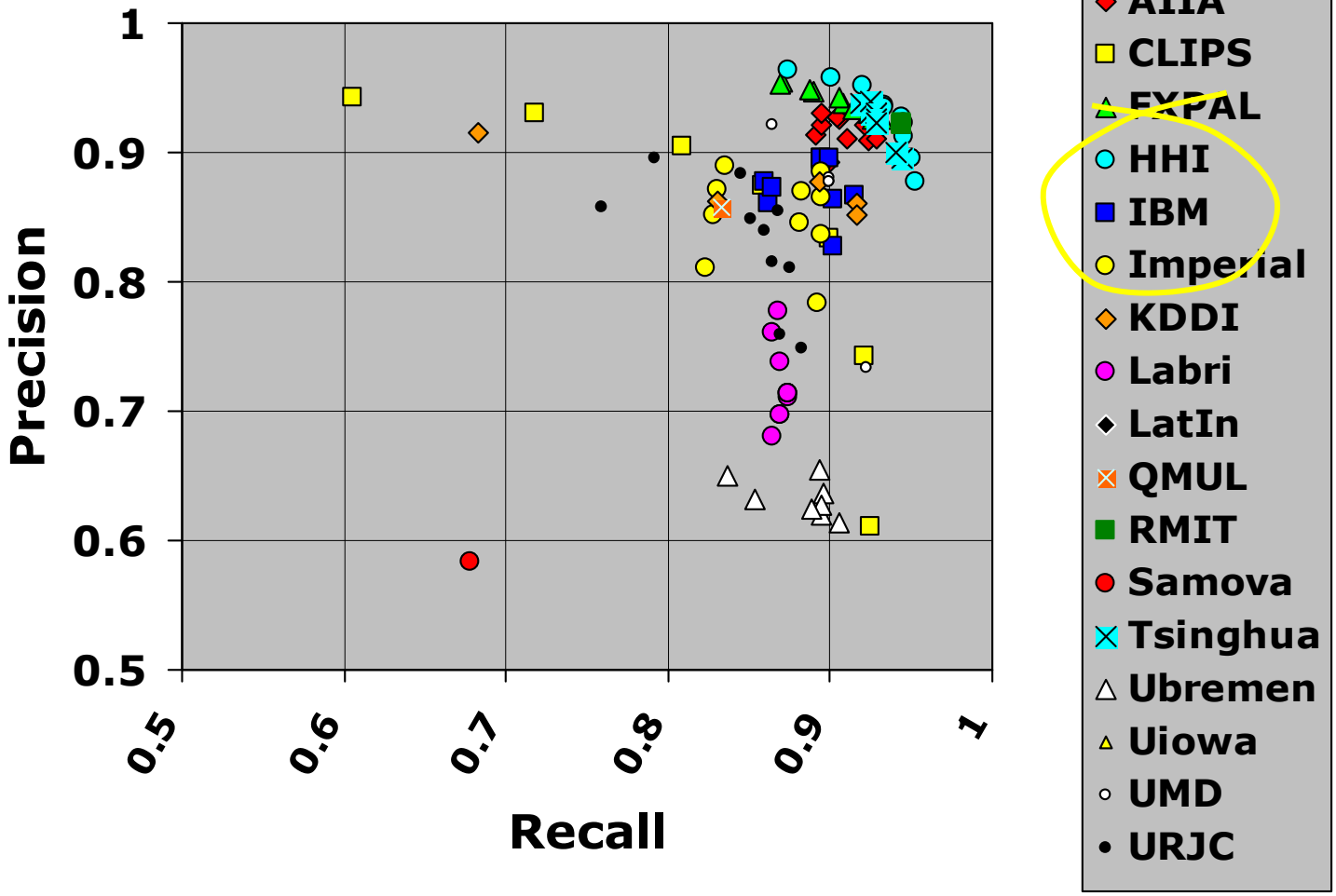
2004: Gradual transitions (zoomed)



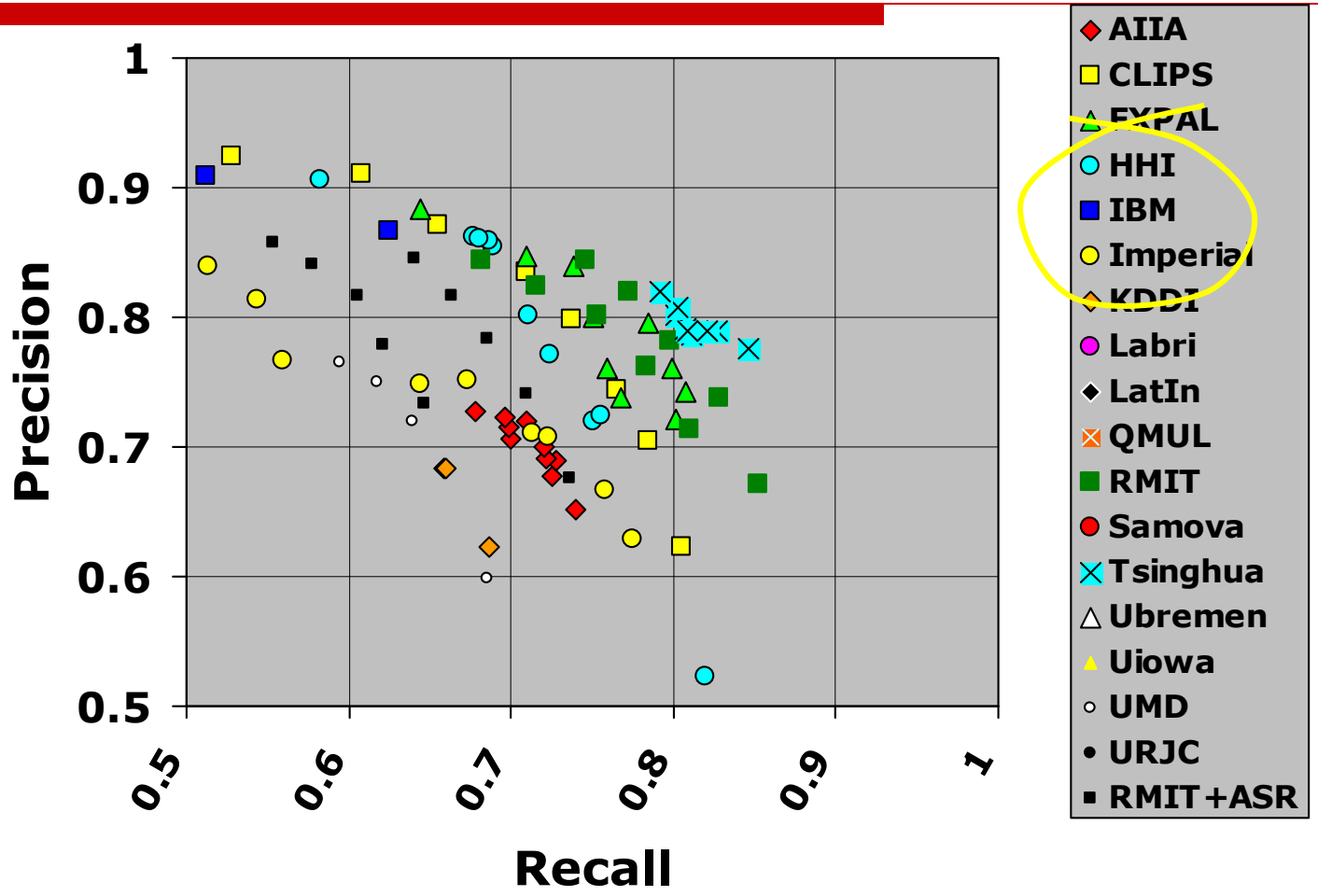
IBM: Headlines

- Approach
  - Same CueVideo System as used previously
- Features
- Performance
  - Using an Intel P4 / 3.06 GHz, runs took 6k to 18k seconds, 3.6 seconds spent on decoding, rest on segmentation, between 3x faster than, and almost RT;
- Results

2004: Cuts (zoomed)



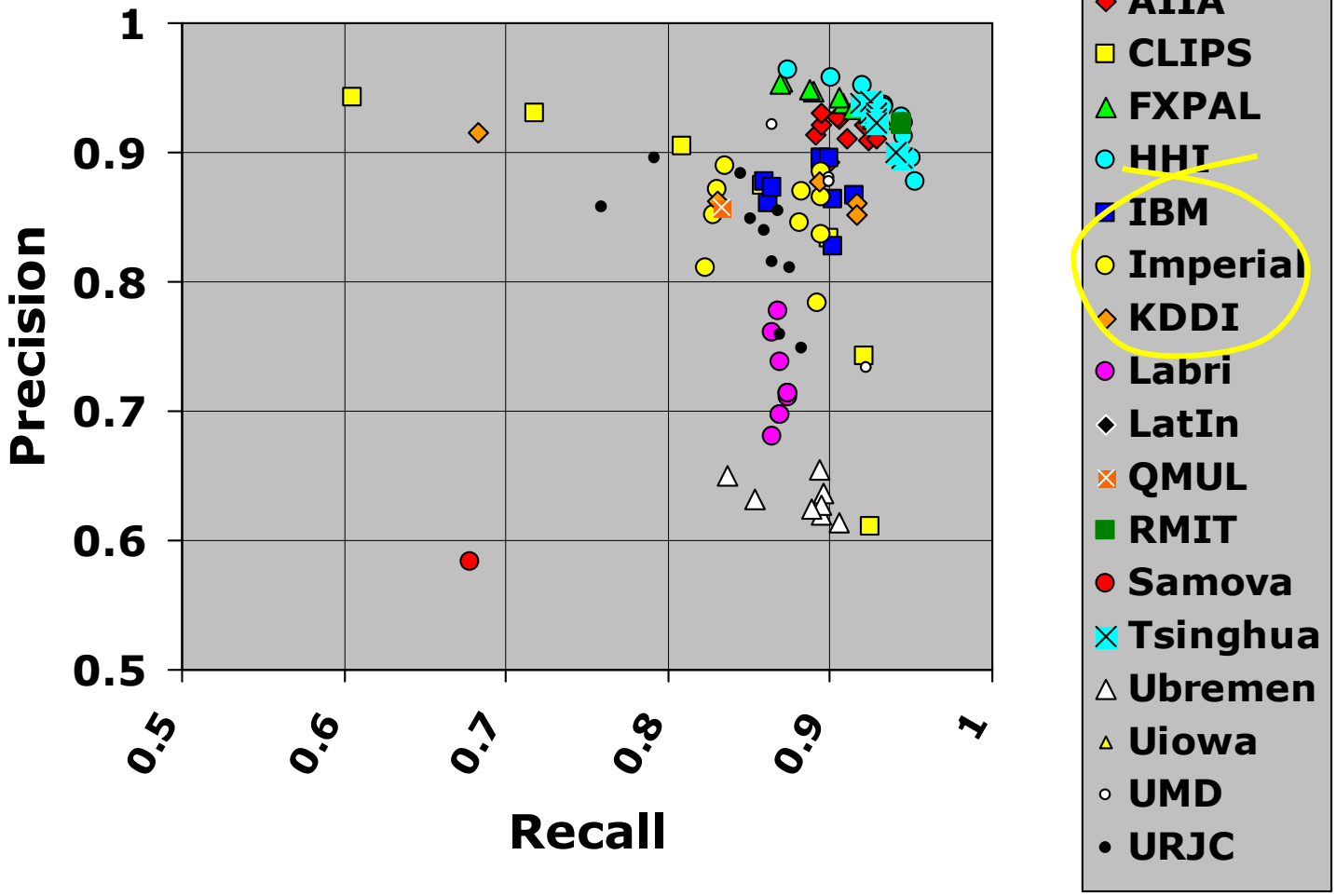
2004: Gradual transitions (zoomed)



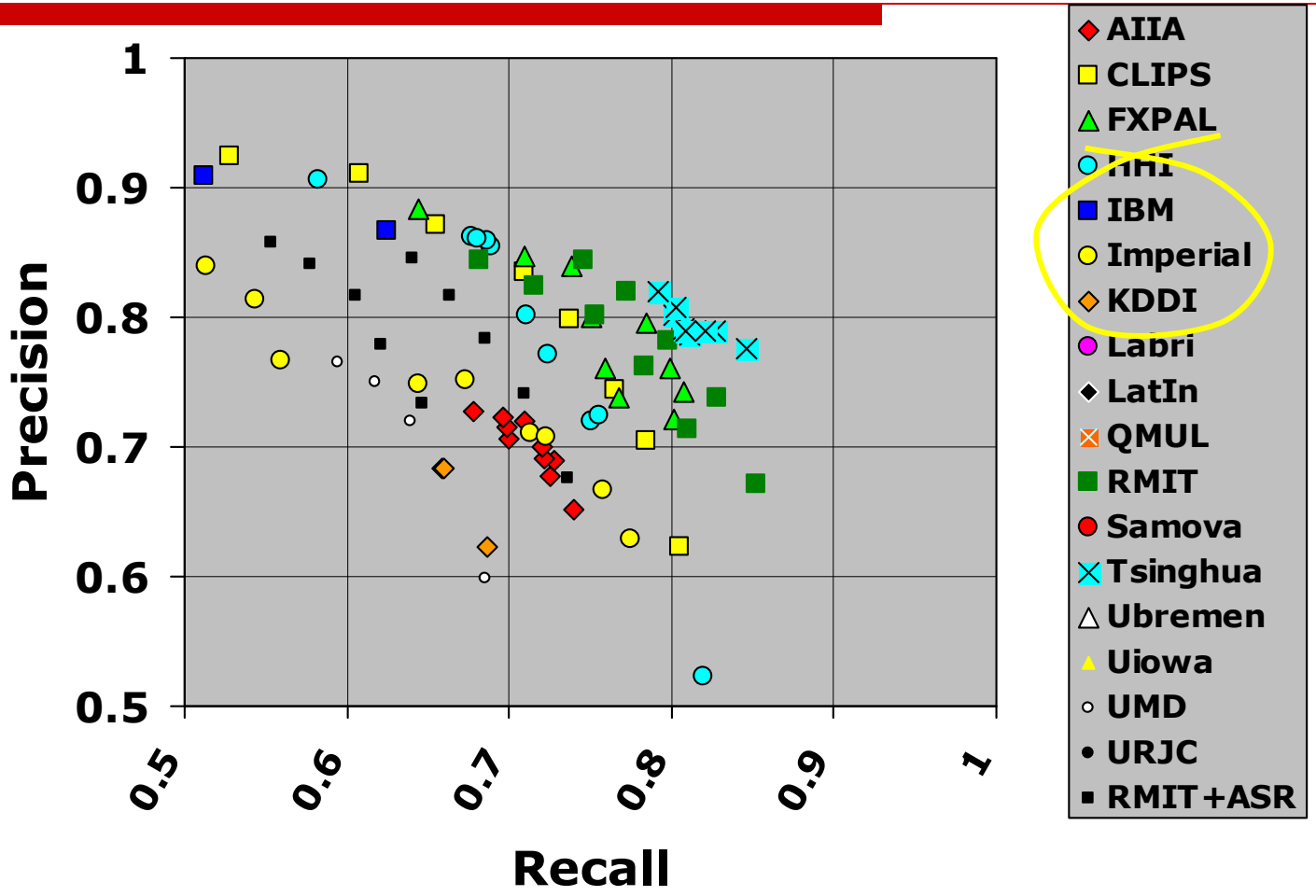
Imperial College London: Headlines

- Approach
  - Distances between colour histograms of frames over a range of timescales;
- Features
  - Largely unchanged from previous years;
- Performance
  - Using an Intel P4 / 2.8 GHz, runs took 9.5k seconds, 2x faster than RT;
- Results

2004: Cuts (zoomed)



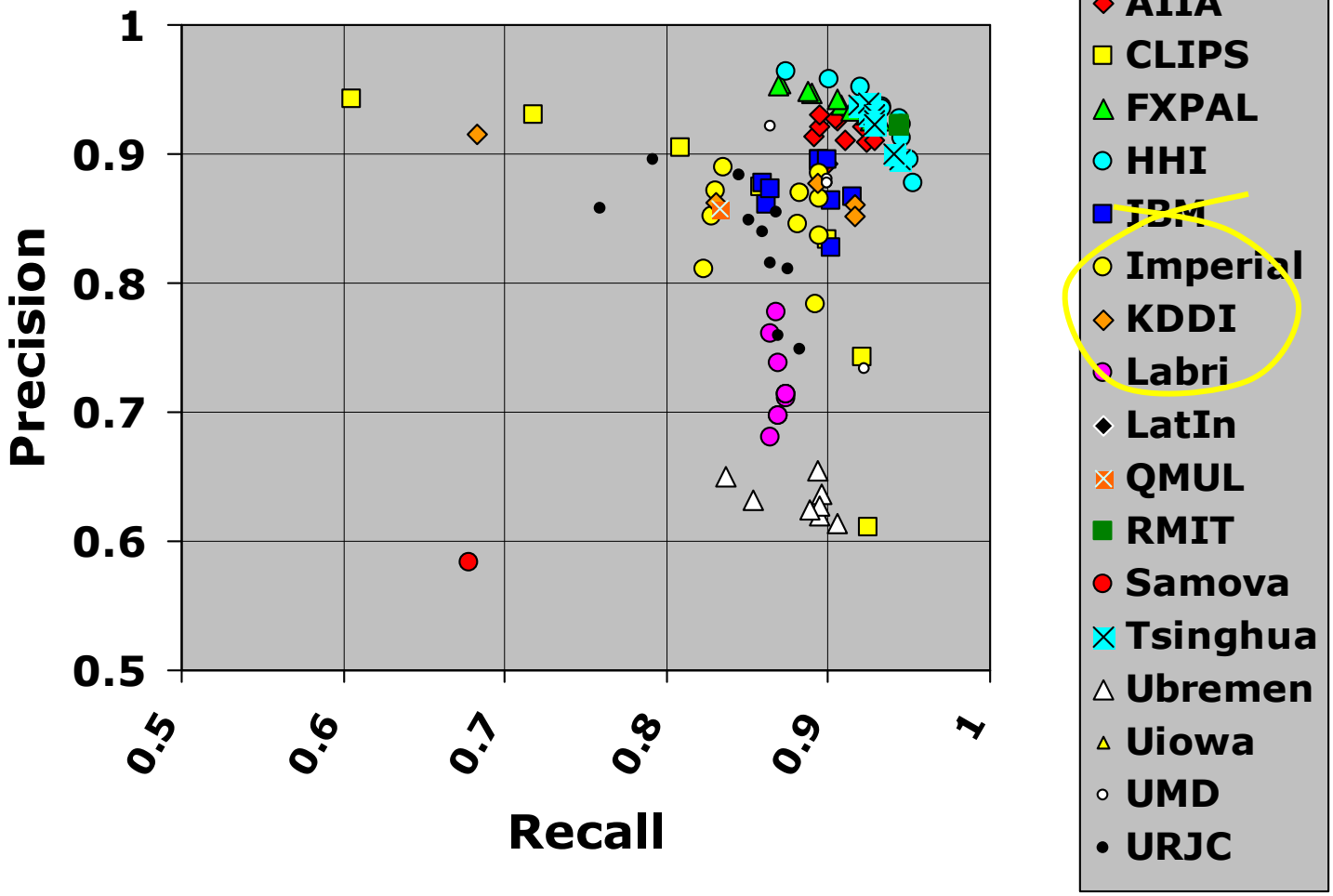
2004: Gradual transitions (zoomed)



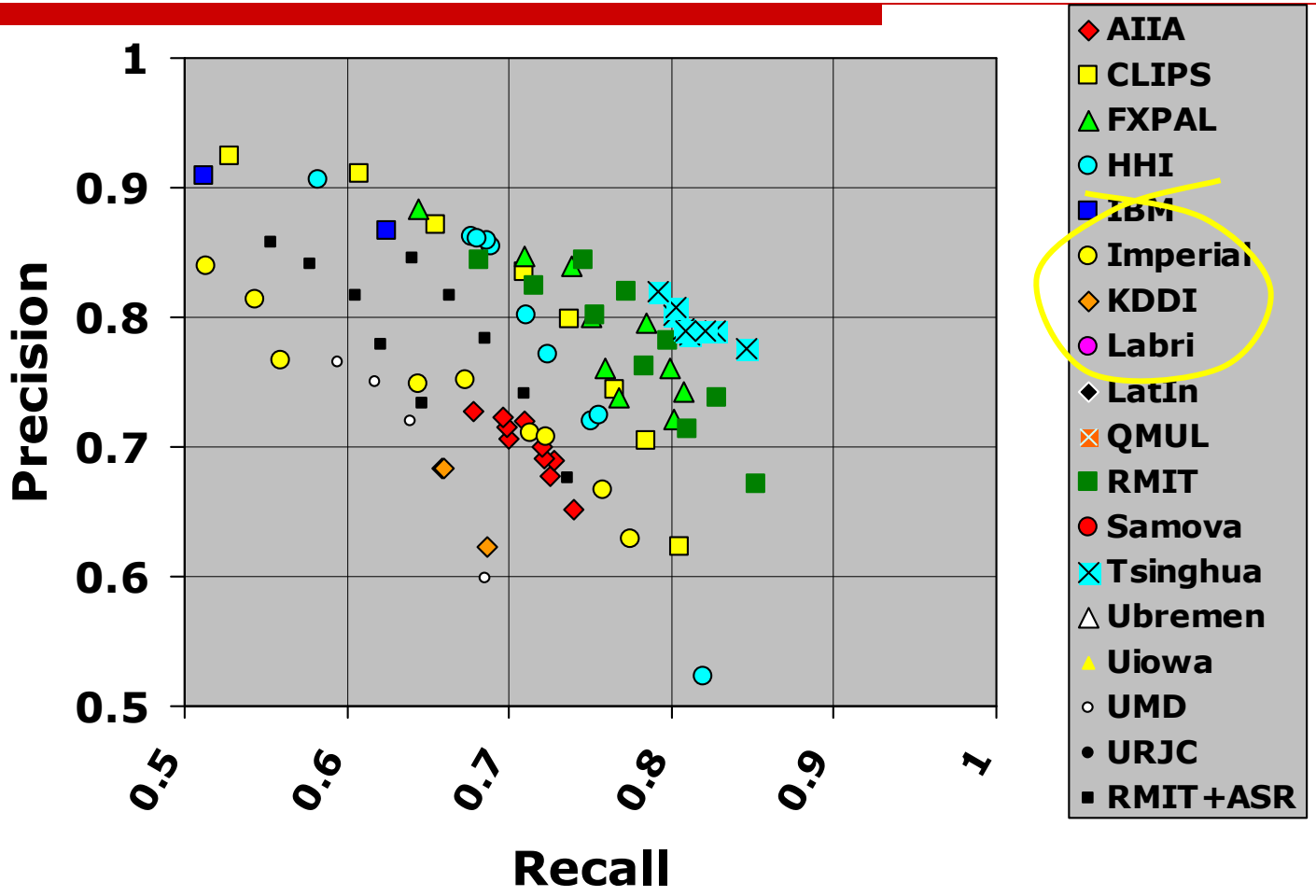
KDDI Labs: Headlines

- Approach
  - Extending the TRECVid2003 approach by adding edge features from DC image, and colour layout, and SVM learning;
  - No series of threshold-varying runs;
  - Adding edge features added a lot of improvement for GTs;
- Features
  - Operates on compressed domain;
- Performance
  - Using an Intel P4 / 1.8 GHz, runs took 0.6k seconds, no decoding, all segmentation, thus very fast - 30x faster than RT ?
- Results

2004: Cuts (zoomed)



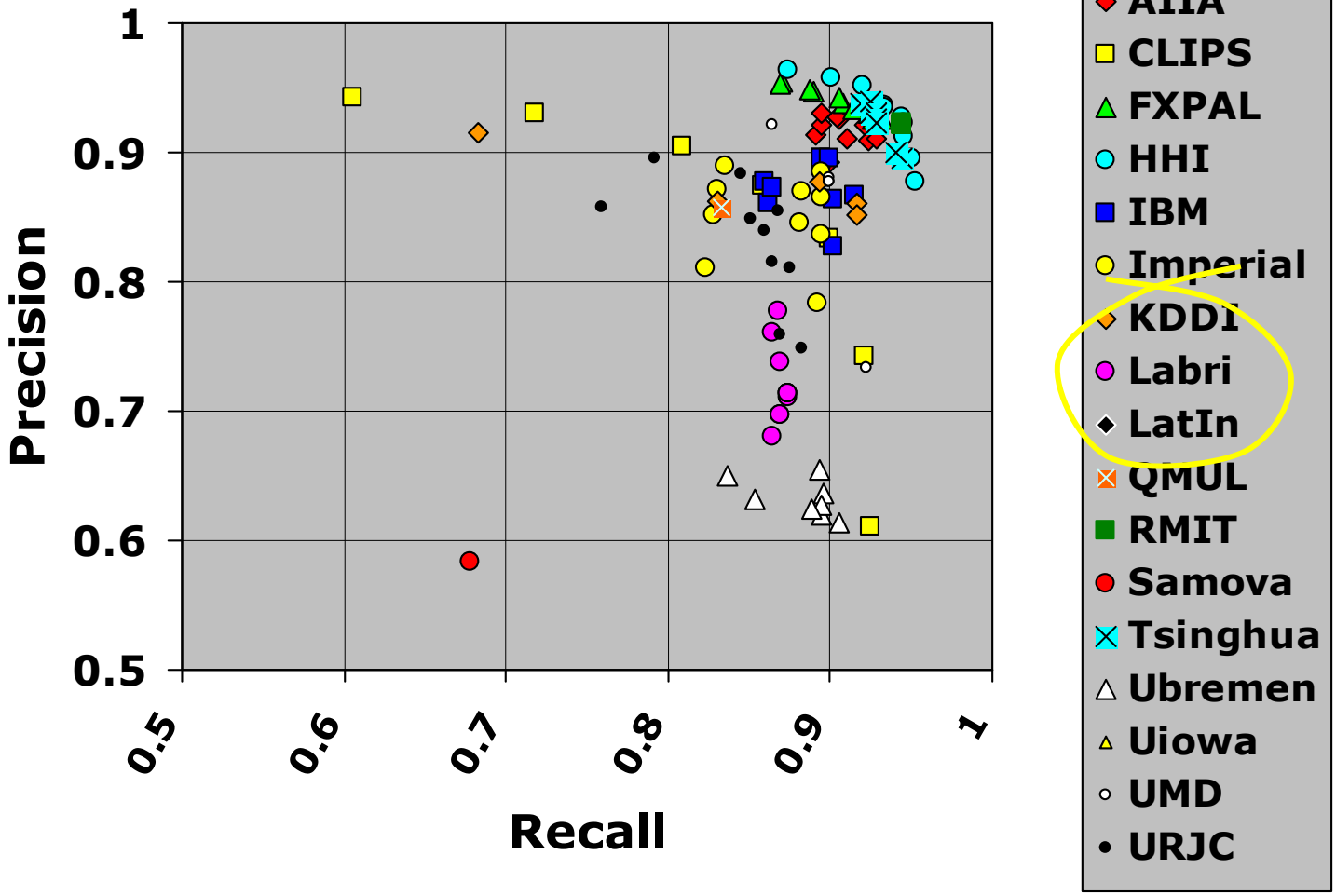
2004: Gradual transitions (zoomed)



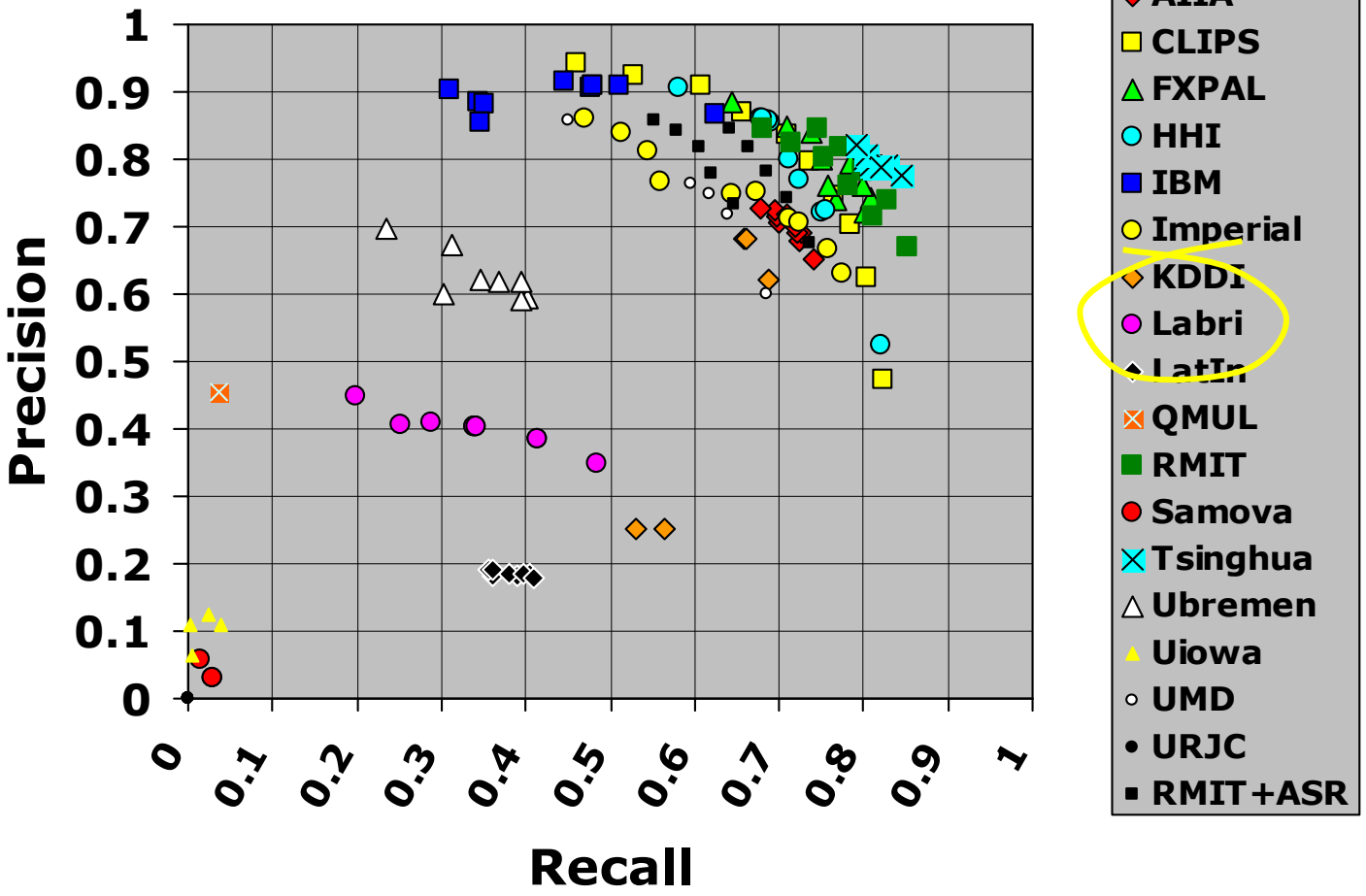
Université Bordeaux - LaBRI: Headlines

- Approach
  - Estimate robust global camera motion
  - Determine P-frame peaks to compute motion and frame statistics, then measure similarity between compensated adjacent I-frames;
- Features
  - Operates on compressed domain, I- and P-frames only
- Performance
  - Using an Intel P4 / 2.8 GHz, runs took 6.7k seconds, 3x faster than RT;
- Results

2004: Cuts (zoomed)



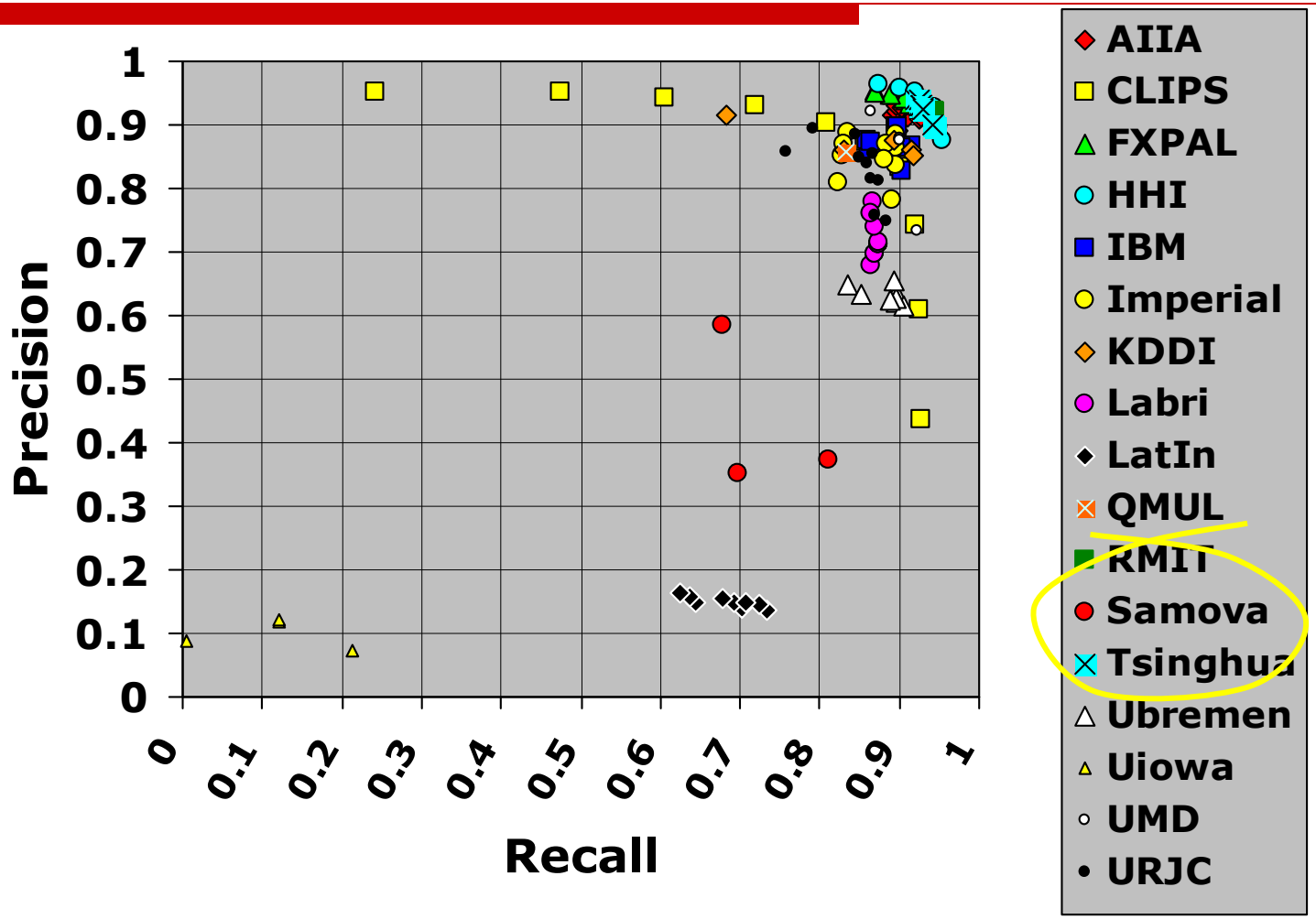
2004: Gradual transitions



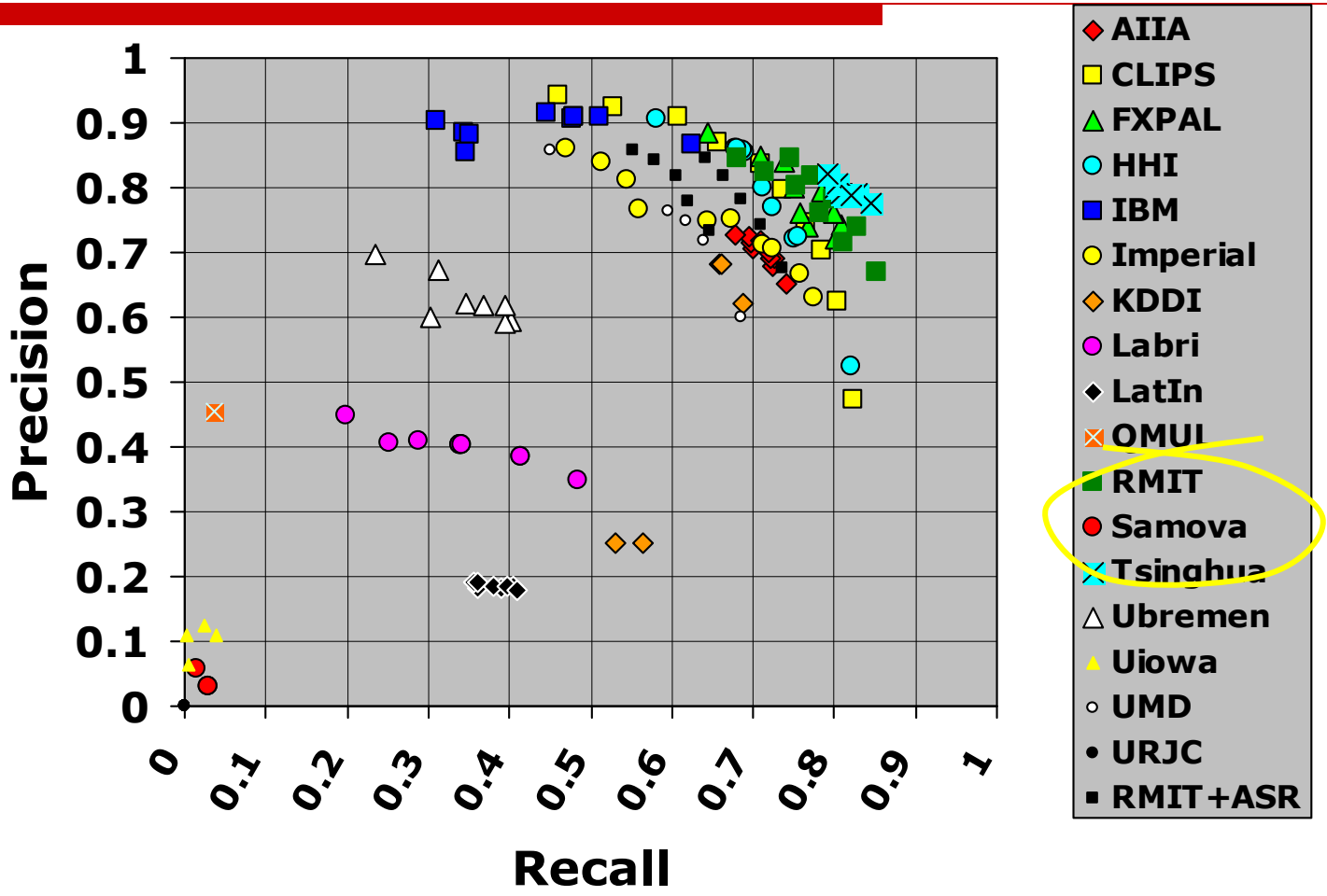
SAMOVA - IRIT Toulouse: Headlines

- Approach
  - Based upon detection and tracking of changes in illumination
- Features
  - Reduce frame size to 44x30 pixels
- Performance
  - Using an Intel P4 / 1.7 GHz, runs took 5.3k seconds, about 4x faster than RT;
- Results

2004: Cuts


Precision
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Recall
AIIA
CLIPS
FXPAL
HHI
IBM
Imperial
KDDI
Labri
LatIn
QMUL
RMIT
Samova
Tsinghua
Ubremen
Uiowa
UMD
URJC

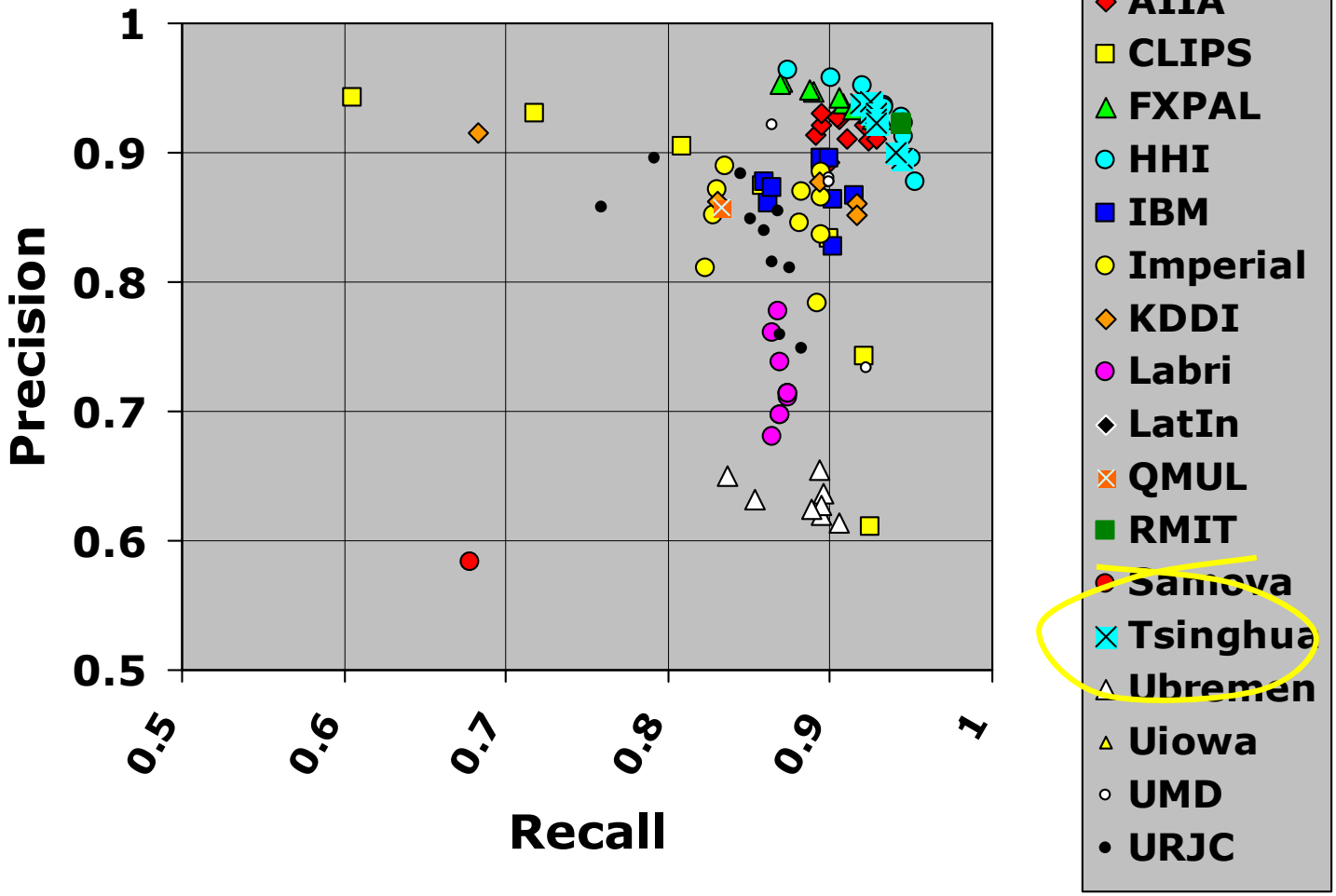
2004: Gradual transitions



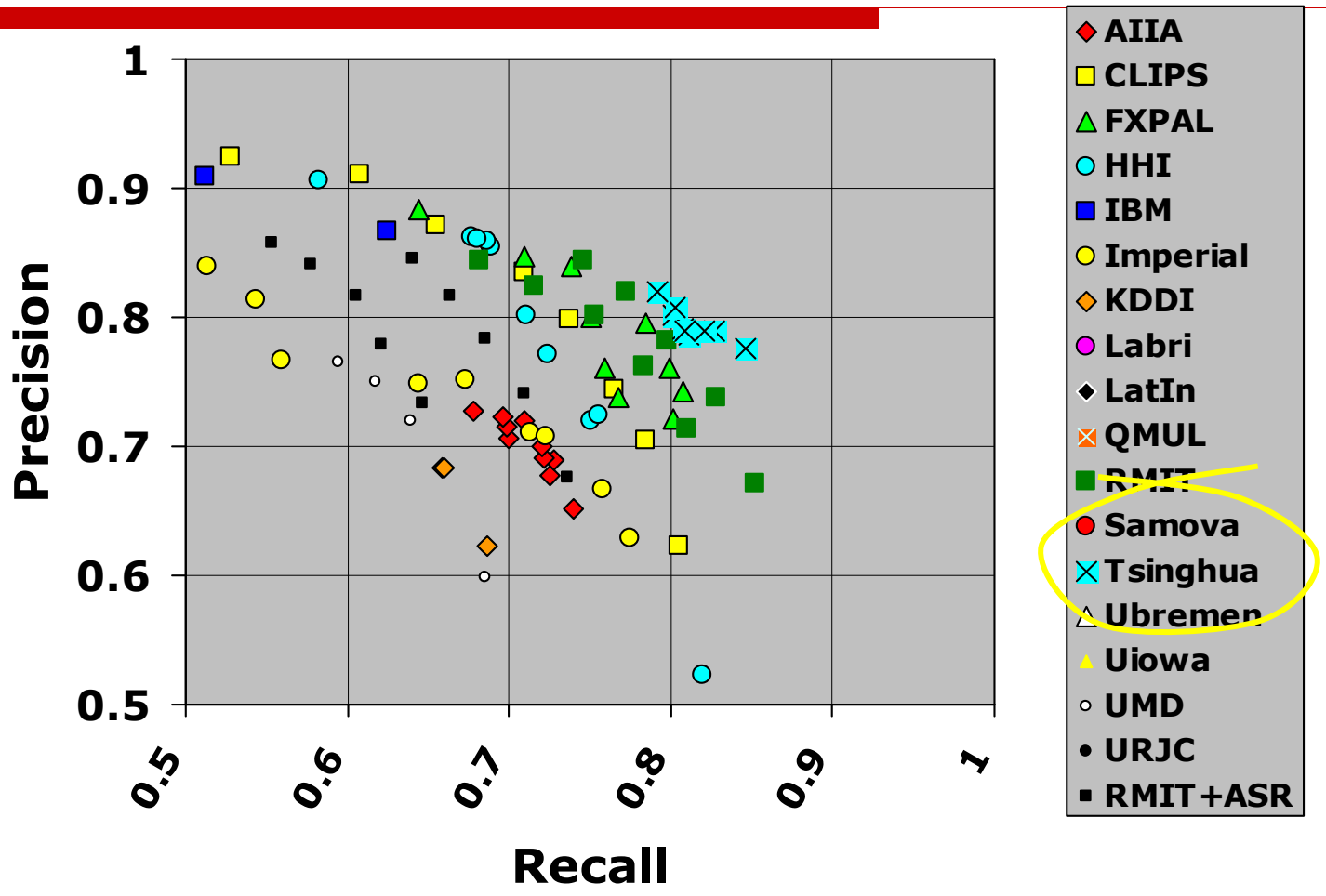
Tsinghua University: Headlines

- Approach
  - A fade in/out detector based on detecting monochrome frames;
  - A cut detector using 2nd order derivatives of colour histogram and pixelwise comparisons, flash detection, and GT elimination;
  - A GT detector using same as cuts, plus motion vectors;
- Features
  - Extension of 2003 work
- Performance
  - Using an AMD 2500+ / 1.84 GHz, runs took 47k seconds, almost all on decoding, +2x real time;
- Results

2004: Cuts (zoomed)



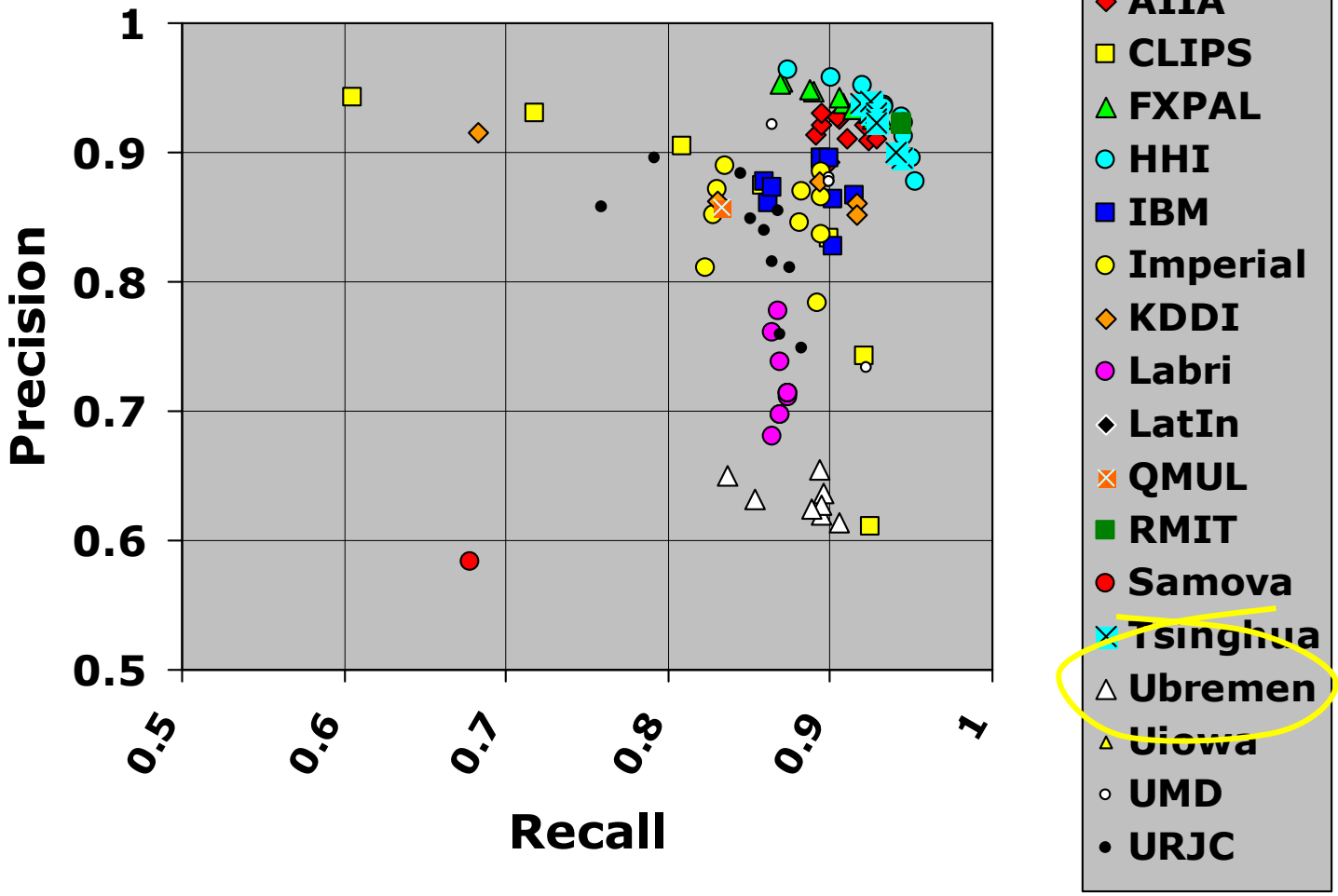
2004: Gradual transitions (zoomed)



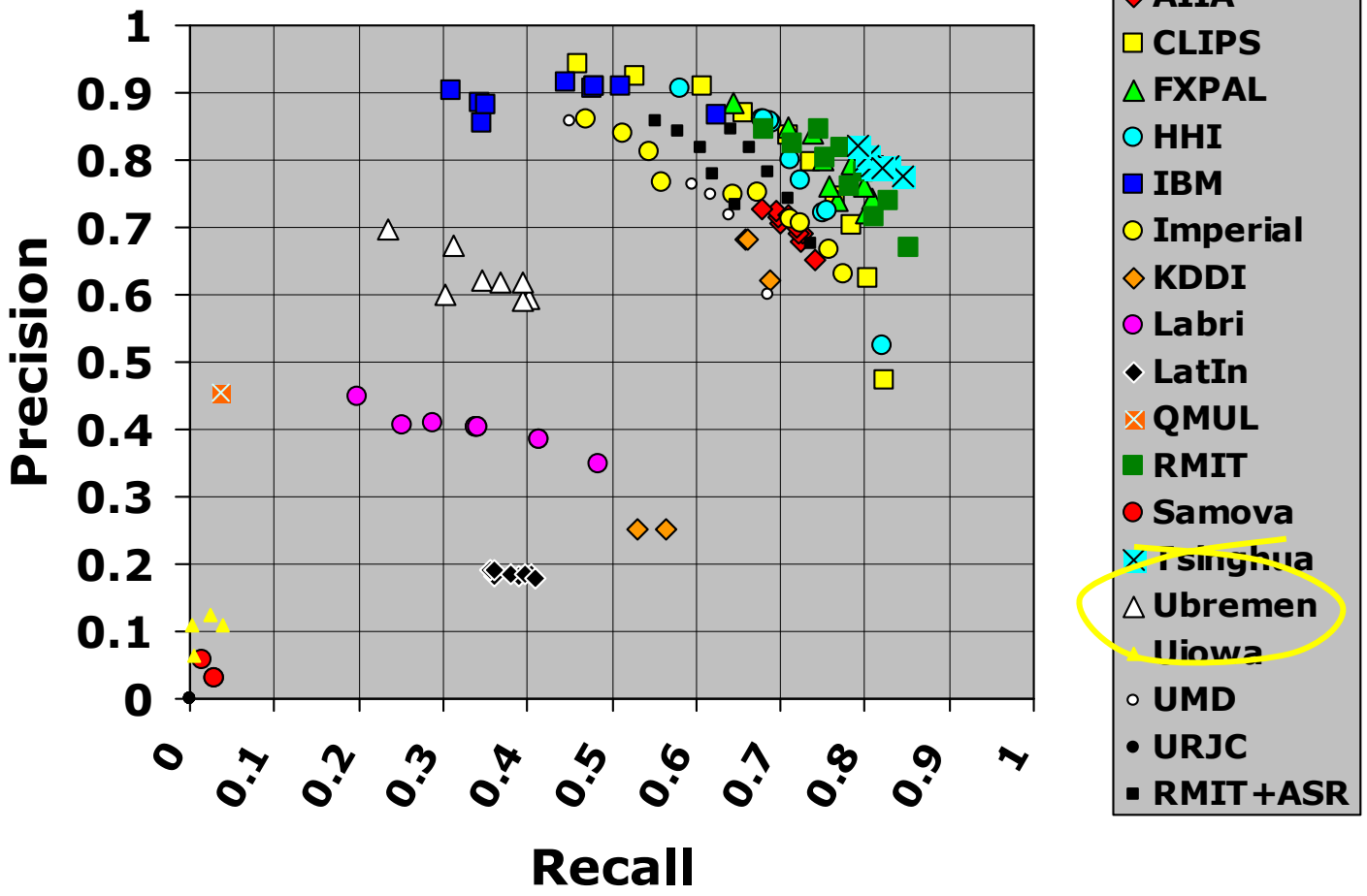
TZI Bremen: Headlines

- Approach
  - RGB histogram differences within 5-frame window, and edge change ration between consecutive frames and in a 10-frame window, followed by block-based motion analysis;
- Features
  - GTs use similar approach to cut detection, but with a FSA
- Performance
  - Using an Intel P4 / 1.7 GHz, runs took 18.3k seconds, just faster than RT
- Results

2004: Cuts (zoomed)



2004: Gradual transitions



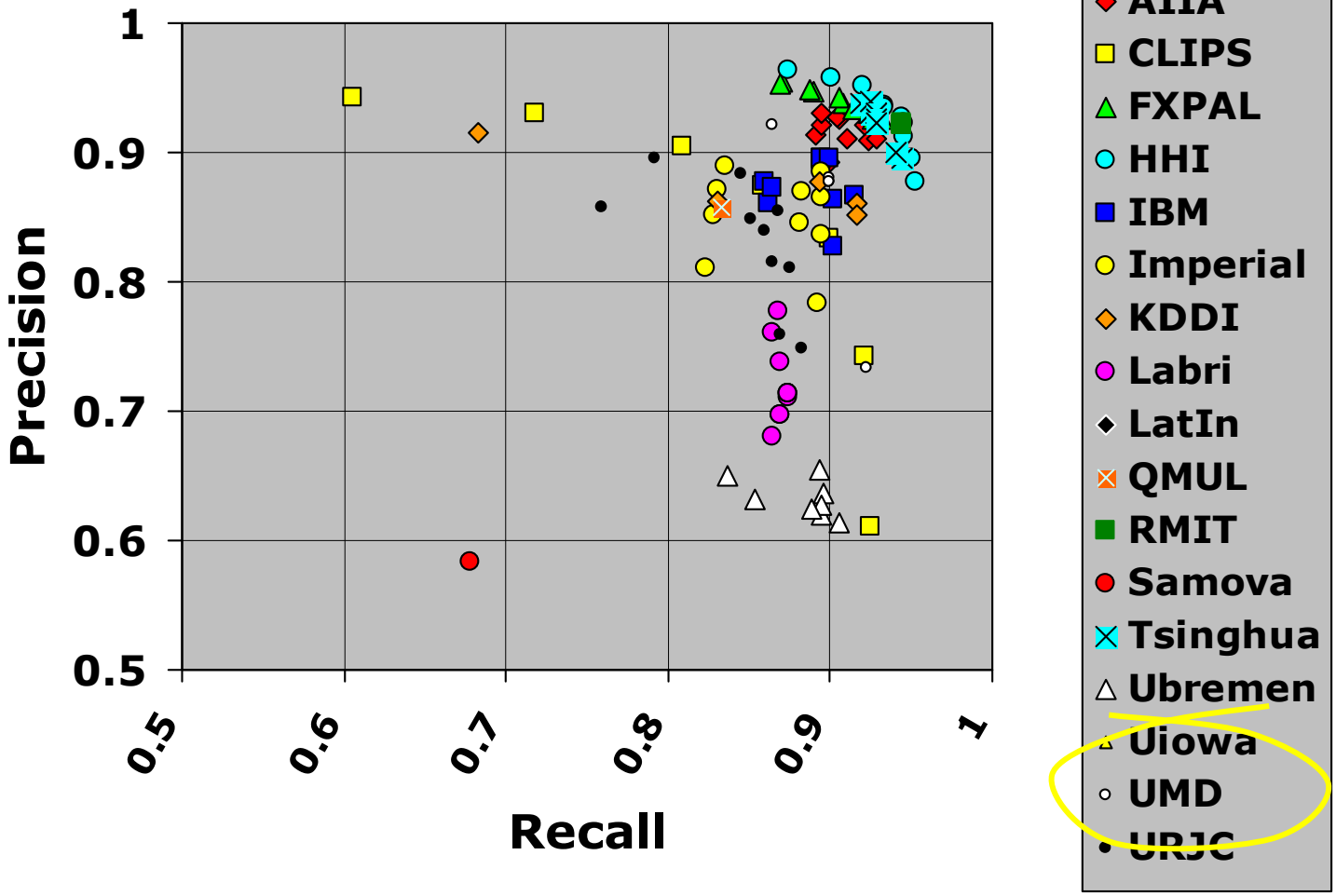
Univ. Iowa: Headlines

- Approach
  - Combination of colour histogram similarity, aggregate colour distance for equivalent pixel pairs, and edge distances;
- Features
  - “technical” faux-pas on submission
- Performance
  - Using an 800 MHz iMacs, runs took 11.5k seconds, almost twice as fast as RT;
- Results

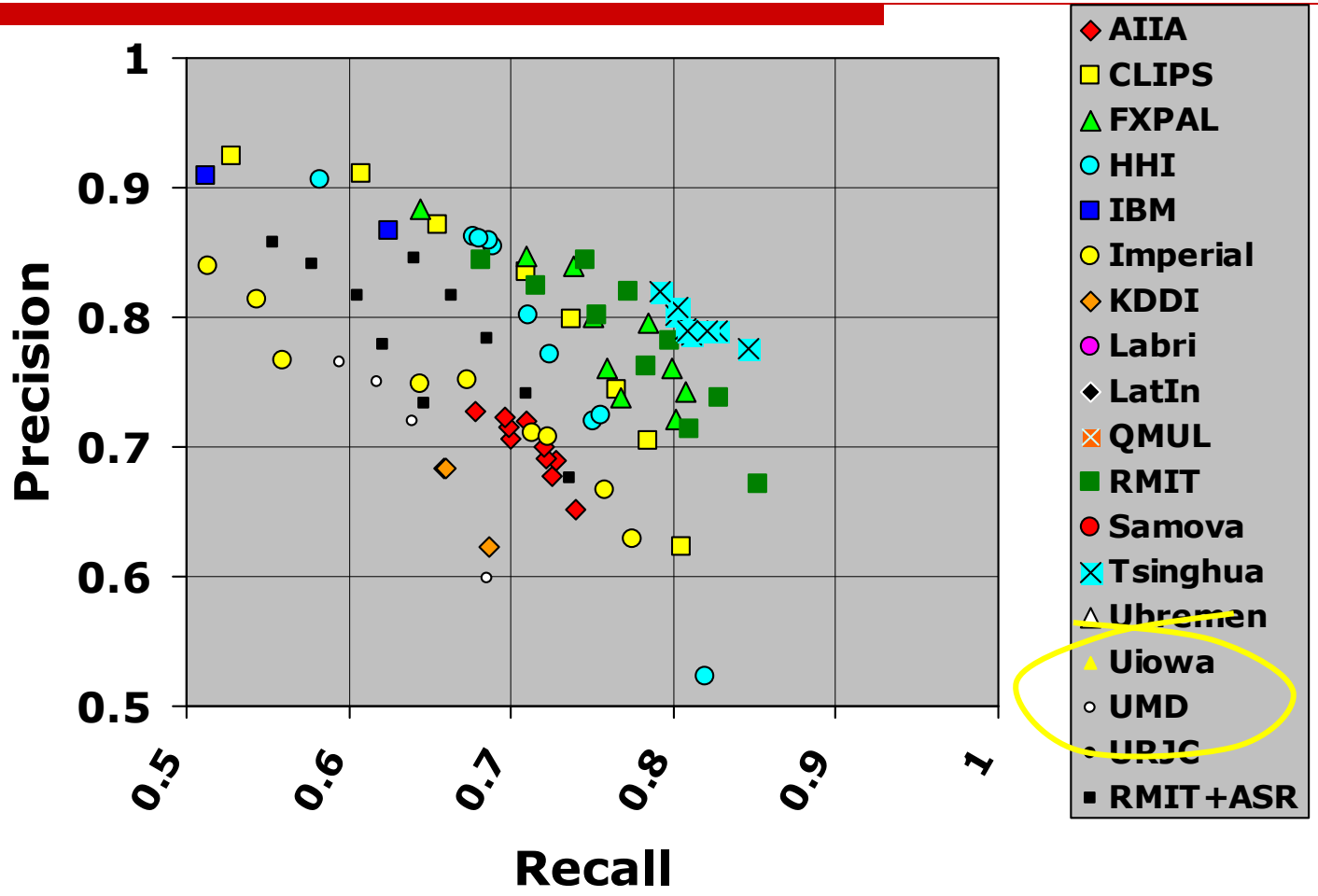
University of Maryland: Headlines

- Approach
  - Not the usual colour histogram or edge differences but 2max ratio
  - For dissolves, uses the skipping image distance, a function of similarity between “distant” frames;
- Features
  - Fast decoding; some misunderstandings on cut/GT categories;
- Performance
  - Using an Intel P4 / 1.8 GHz, runs took 11k seconds, nearly 2x faster than RT;
- Results

2004: Cuts (zoomed)



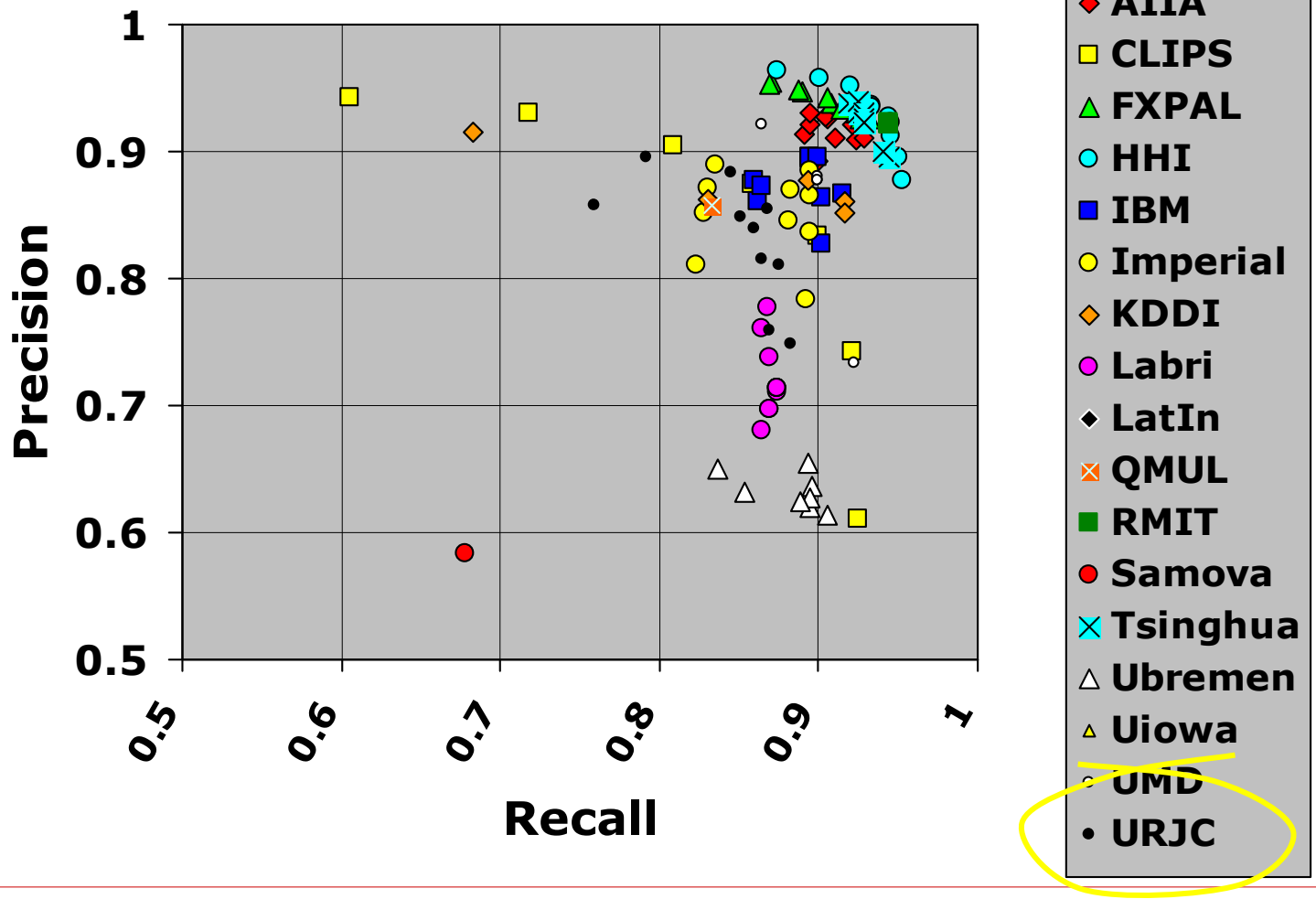
2004: Gradual transitions (zoomed)



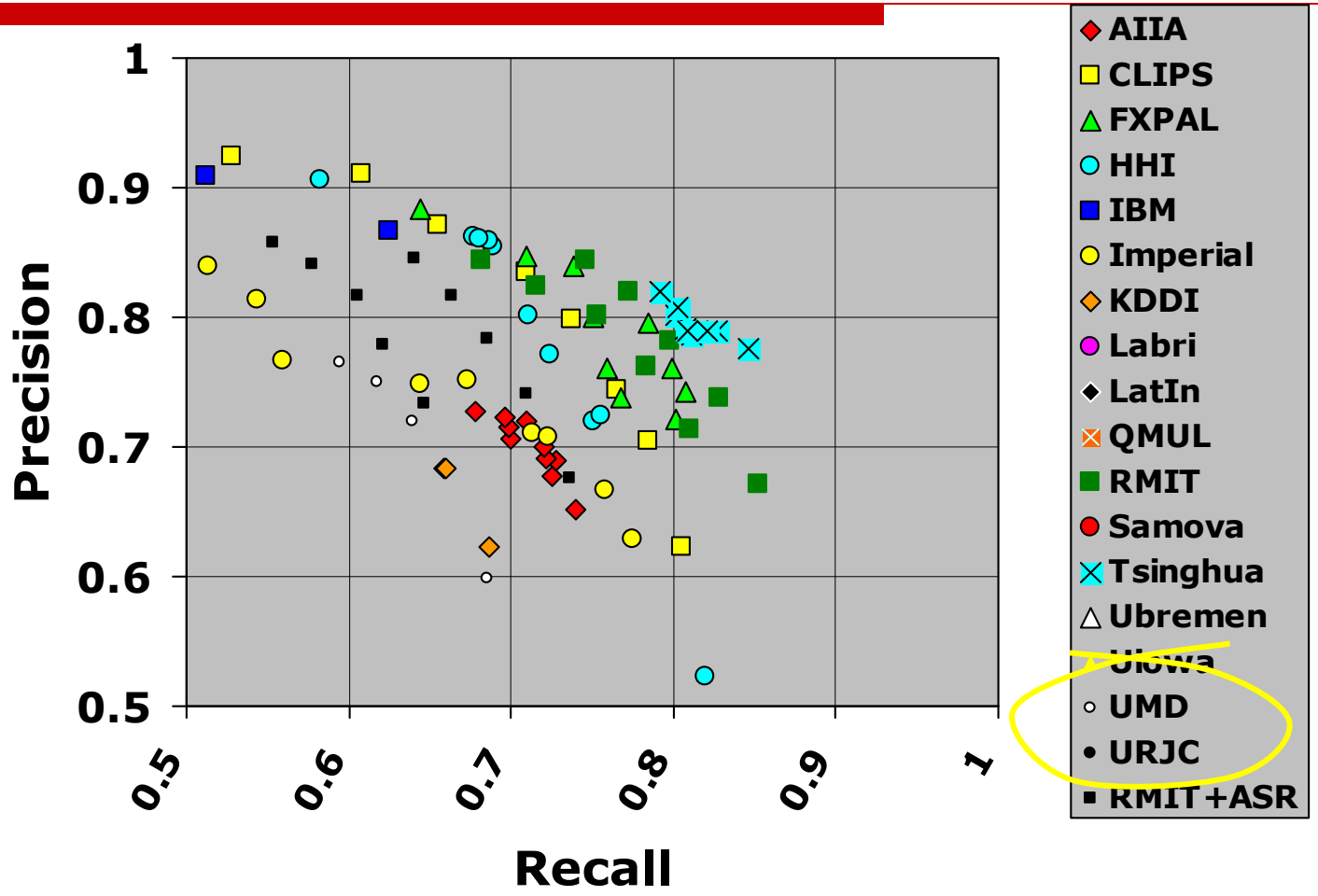
U. Rey Juan Carlos (Madrid): Headlines

- Approach
  - Colour histogram differences, different bin sizes;
- Features
  - Redistribution of boundary values yielding multiresolution histograms;
- Performance
  - Using an Intel P4 / 3.0 GHz, runs took 5.5k seconds, nearly 4x faster than RT;
- Results

2004: Cuts (zoomed)



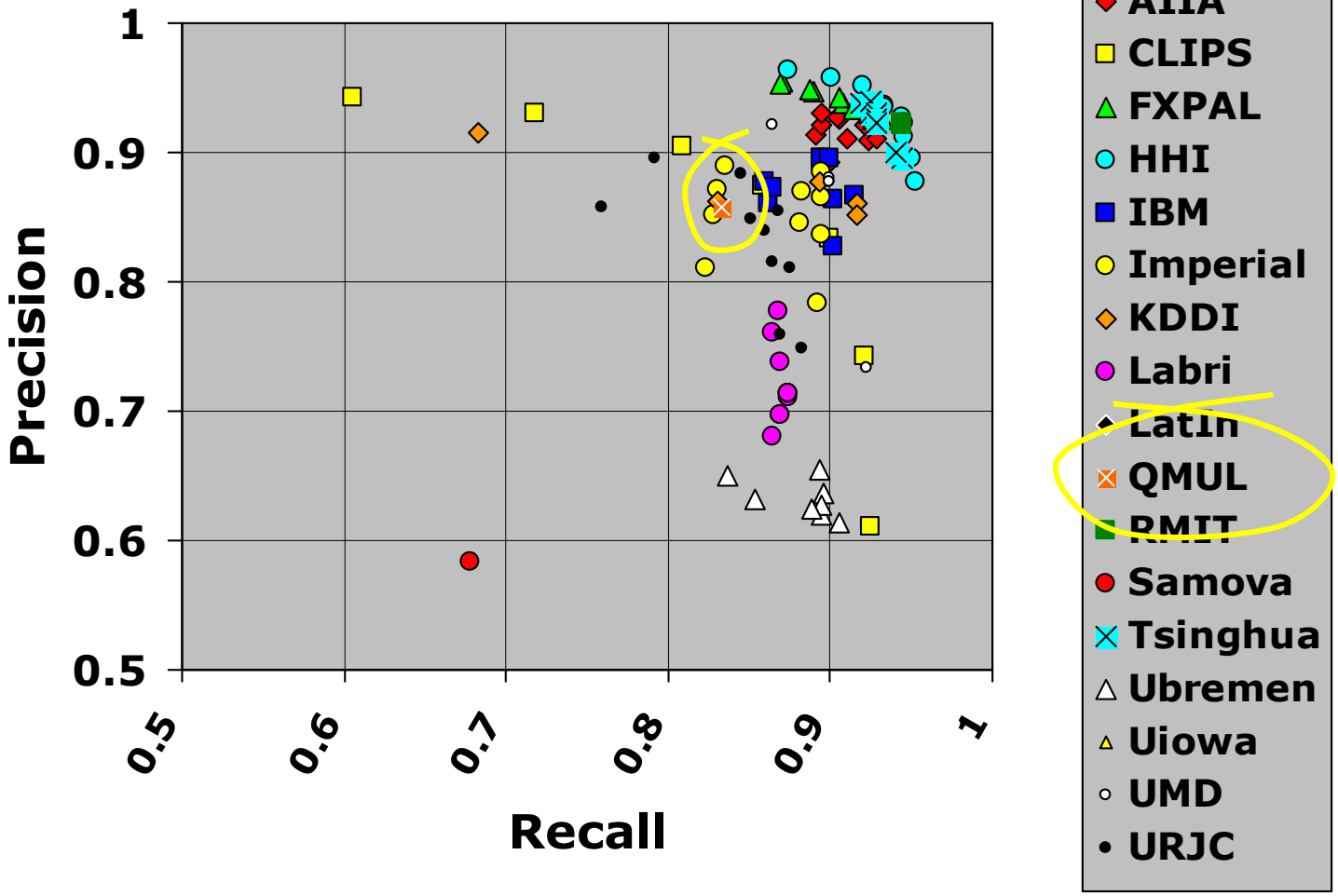
2004: Gradual transitions (zoomed)



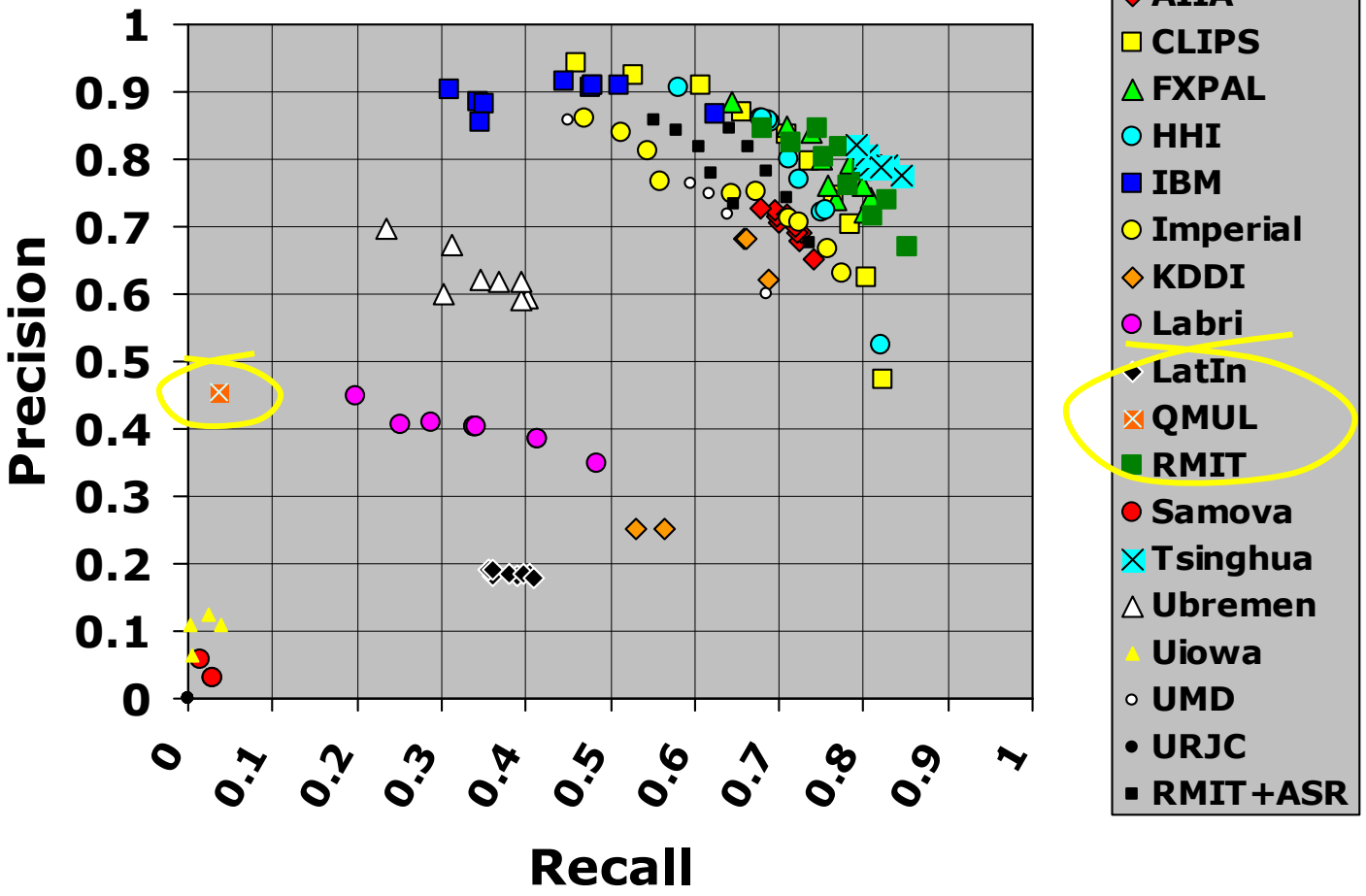
Other Headlines

- No details of submitted runs available yet from QMUL (London), University Sao Paolo (Brazil) or AIIA (Greece), and their results …

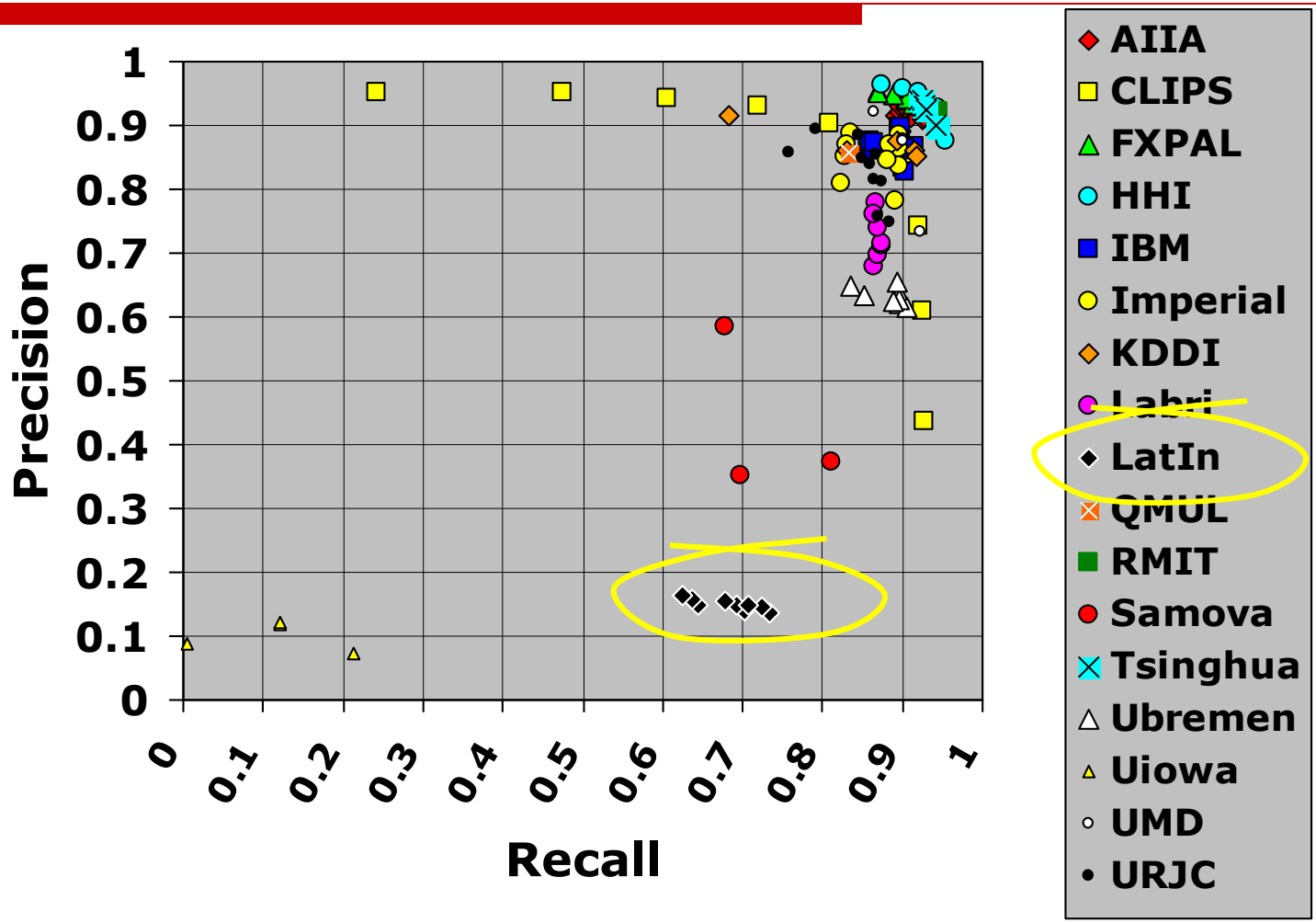
2004: Cuts (zoomed)



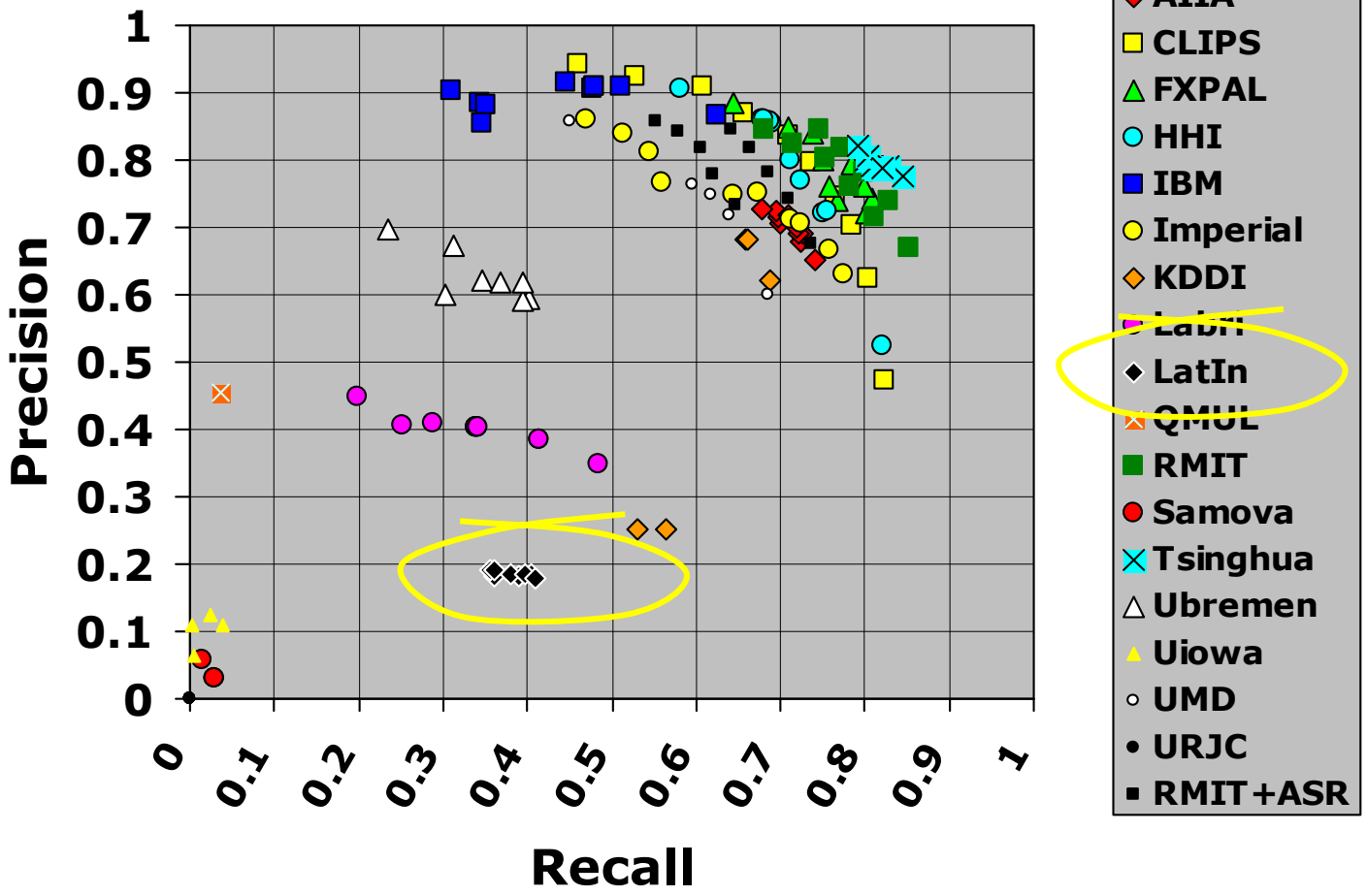
2004: Gradual transitions



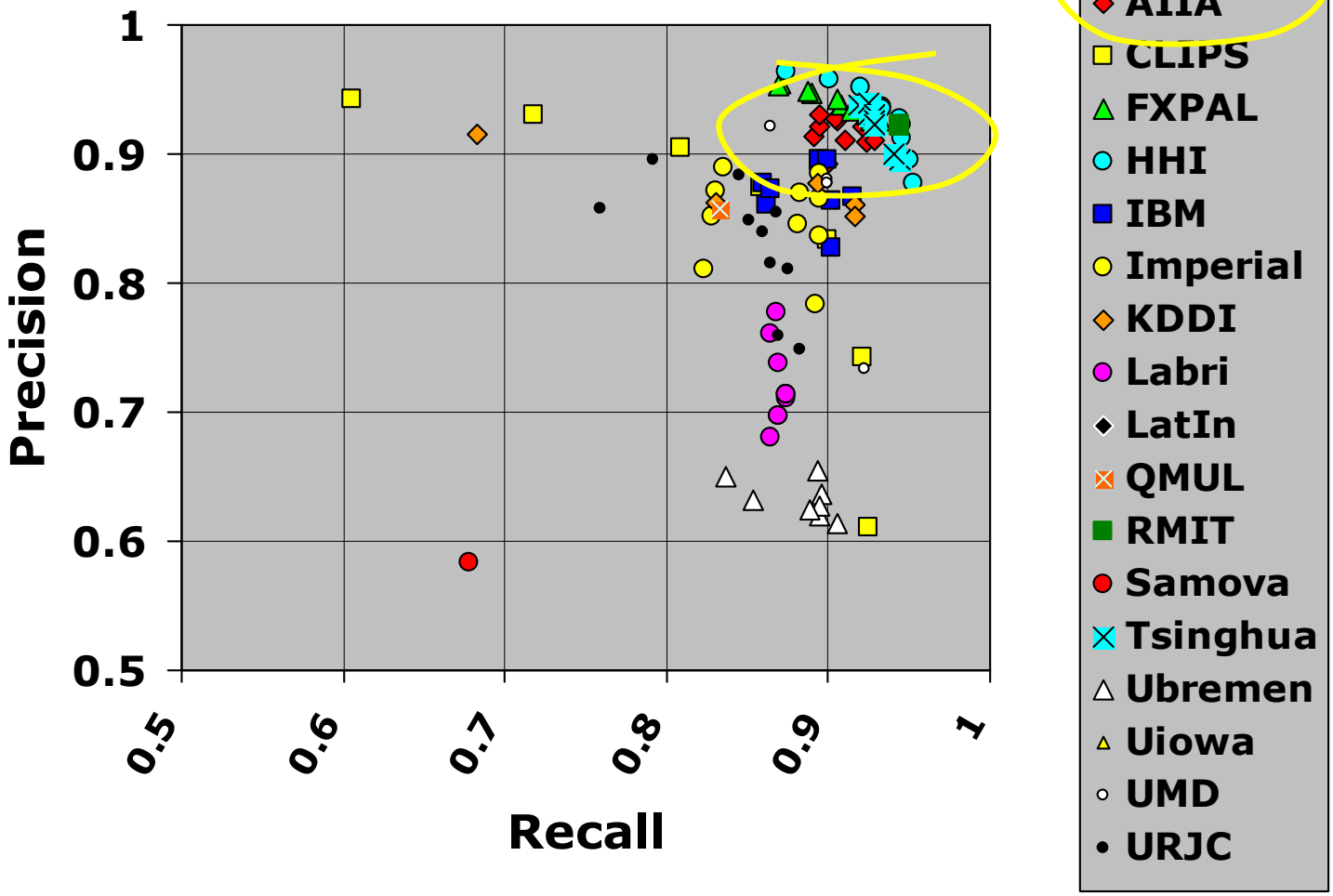
2004: Cuts


Precision
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Recall
AIIA
CLIPS
FXPAL
HHI
IBM
Imperial
KDDI
Labri
LatIn
QMUL
RMIT
Samova
Tsinghua
Ubremen
Uiowa
UMD
URJC

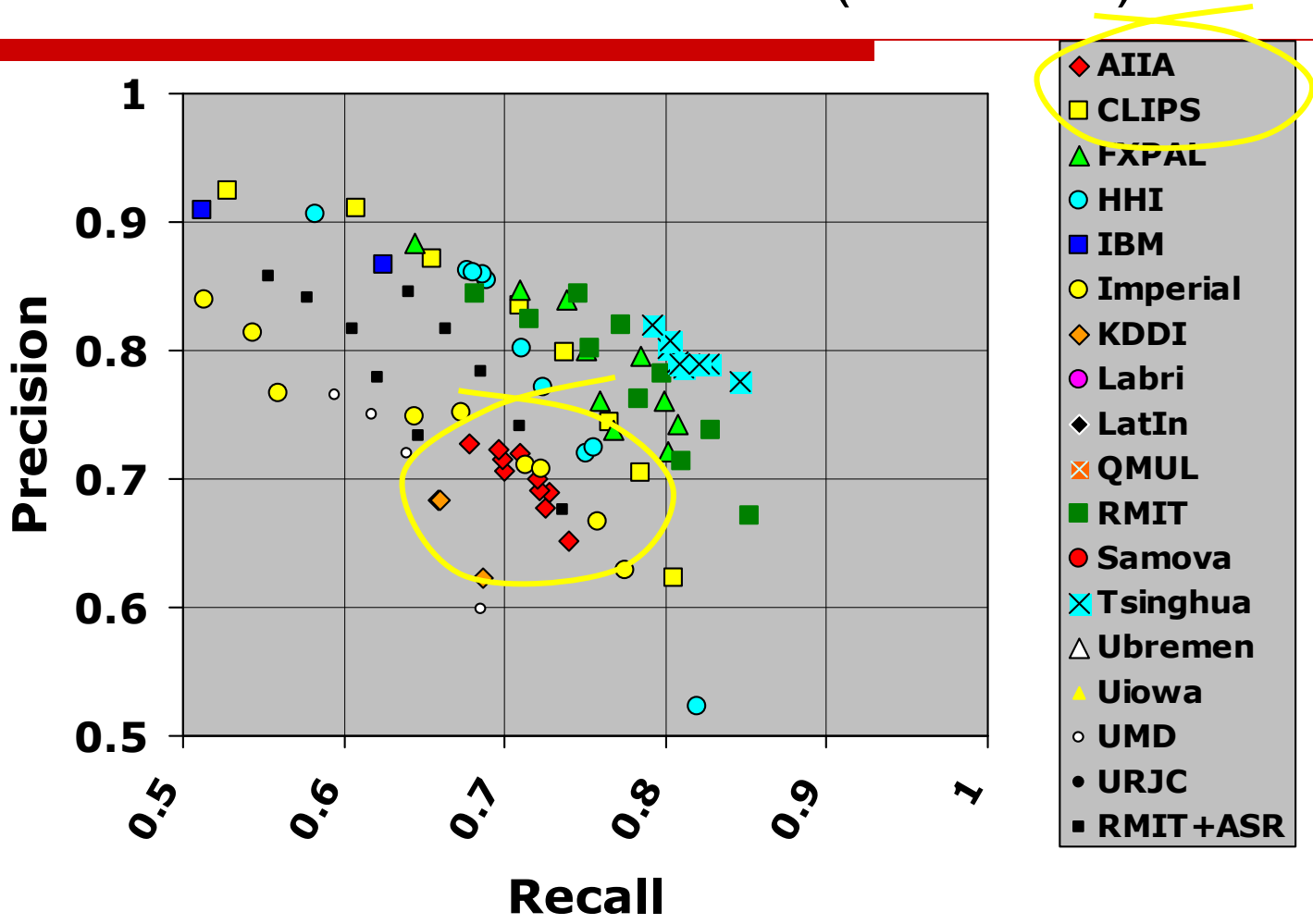
2004: Gradual transitions



2004: Cuts (zoomed)



2004: Gradual transitions (zoomed)



Conclusions …

- Strong interest;
- Novel approaches continue to emerge;
- Adding computation cost was a good idea, and worked well, even though done after the fact, and was voluntary (thanks participants !);
- Lots of data available to do a more comprehensive comparative analysis;