TRECVID-2005:
Shot Boundary Detection Task Overview

Alan Smeaton
Dublin City University
&
Paul Over
NIST

SB Task Definition

- Shot boundary detection is a fundamental task in any kind of video content manipulation
- Task provides a good entry for groups who wish to “break into” video retrieval and TRECVID gradually
- 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
- 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:
  - Total runtime in seconds
    - Total decode time in seconds
    - Total segmentation time in seconds
  - Processor description

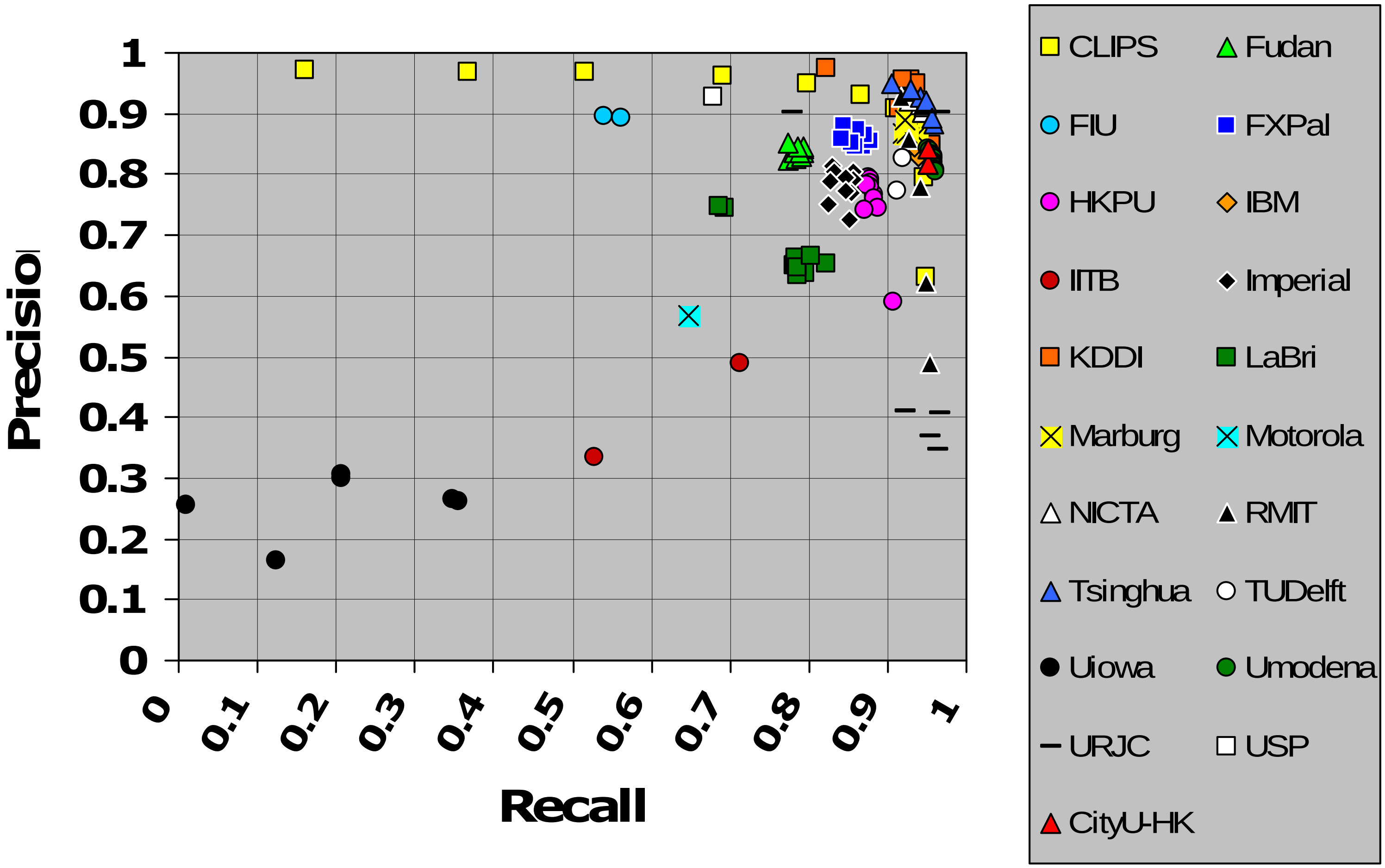
Shot boundary task: Participating groups

City University of Hong Kong	China	SB LL -- --
CLIPS-IMAG, LSR-IMAG, Laboratoire LIS	France	SB -- HL --
Florida International University	USA	SB -- -- --
Fudan University	China	SB LL HL SE
FX Palo Alto Laboratory	USA	SB -- HL SE
Hong Kong Polytechnic University	China	SB -- -- --
IBM	USA	SB -- HL SE
Imperial College London	UK	SB -- HL SE
Indian Institute of Technology (IIT)	India	SB -- -- --
KDDI R\&D Laboratories, Inc.	Japan	SB LL -- --
LaBRI	France	SB LL -- --
Motorola Multimedia Research Laboratory	USA	SB -- -- --
National ICT Australia	Australia	SB LL HL --
RMIT University	Australia	SB -- -- --
Technical University of Delft	Netherlands	SB -- -- --
Tsinghua University	China	SB LL HL SE
University of Central Florida / University of Modena	USA,Italy	SB LL HL SE
University of Iowa	USA	SB LL -- SE
University of Marburg	Germany	SB LL -- --
University Rey Juan Carlos	Spain	SB -- -- --
University of Sao Paulo (USP)	Brazil	SB -- -- --

Shot boundary data

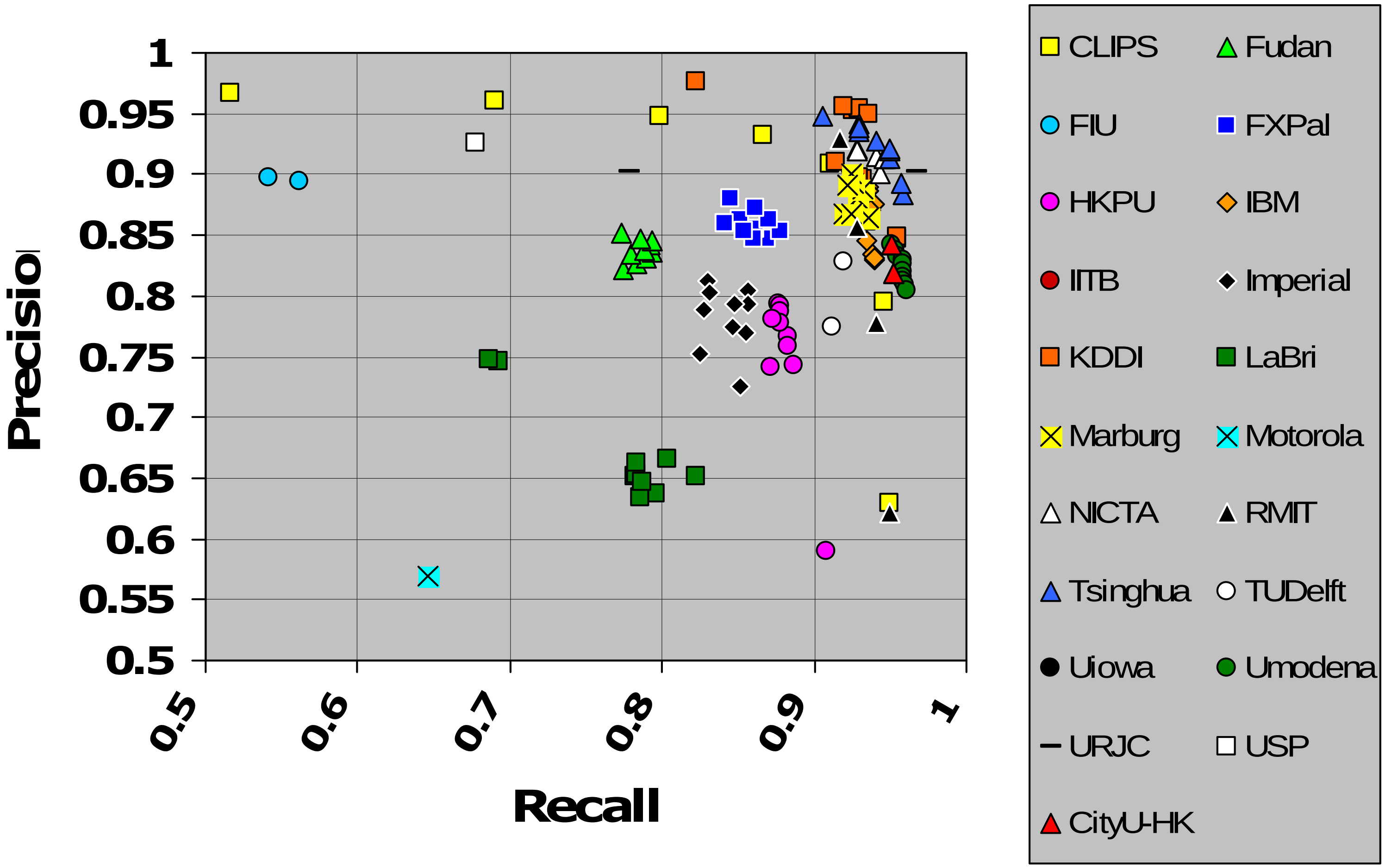
- 12 representative videos (8 news, 4 NASA)
- Total frames: 744,604
- Total transitions: 4,535
- 0.609 transitions/100frames (down from 0.777 in 2004)
- Transition types:
  - **2,759 (60.8%) Cuts** (2004: 57.7%)
  - **1,382 (30.5%) Dissolves** (2004:31.7%)
  - **81 (1.8%) Fade-out/-in** (2004: 4.8%)
  - **313 (6.9%) other** (2004: 5.7%)

Cuts



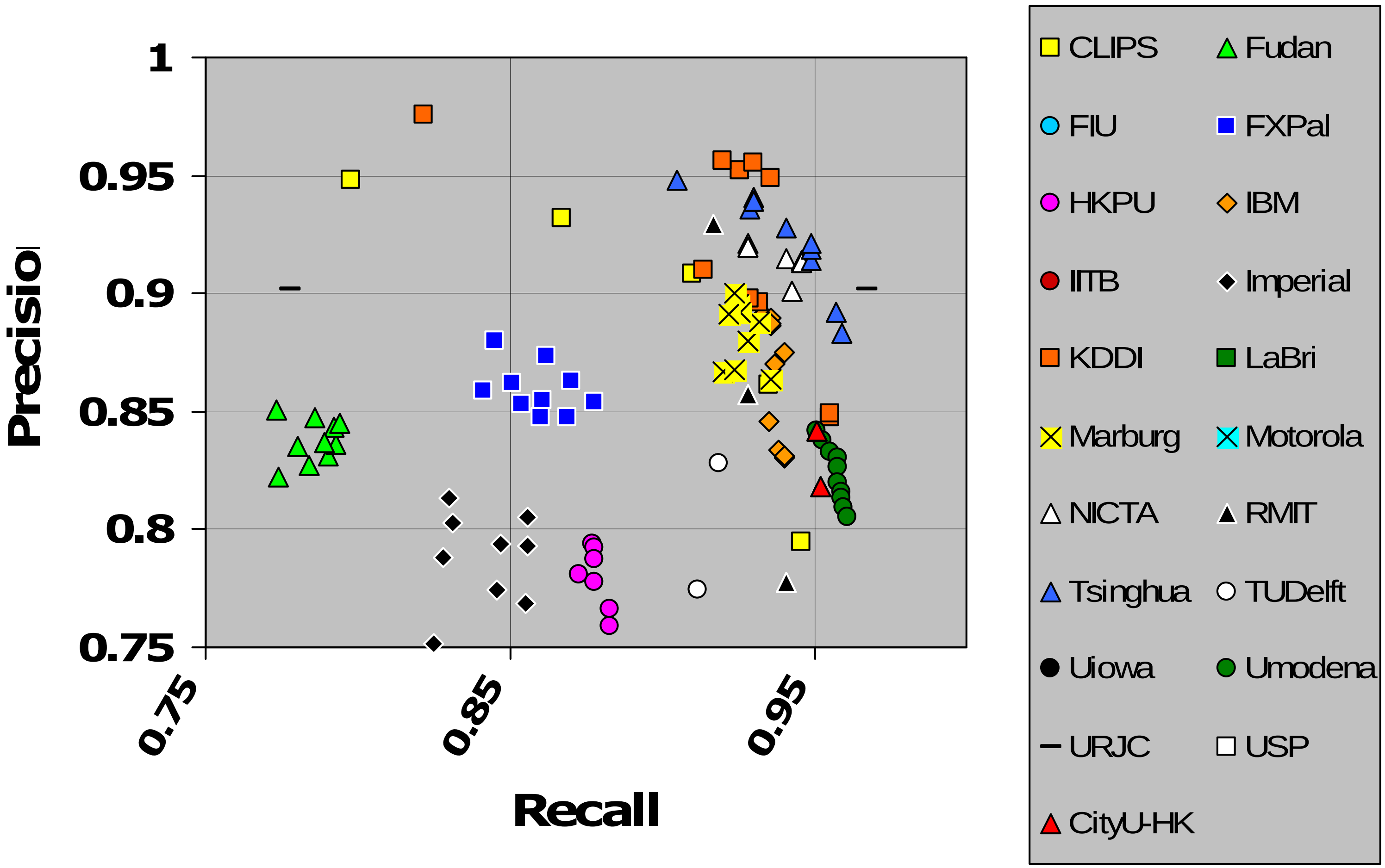
Precisio
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
CLIPS
Fudan
FIU
FXPal
HKPU
IBM
IITB
Imperial
KDDI
LaBri
Marburg
Motorola
NICTA
RMIT
Tsinghua
TUDelft
UIowa
Umodena
URJC
USP
CityU-HK

Cuts (zoomed)

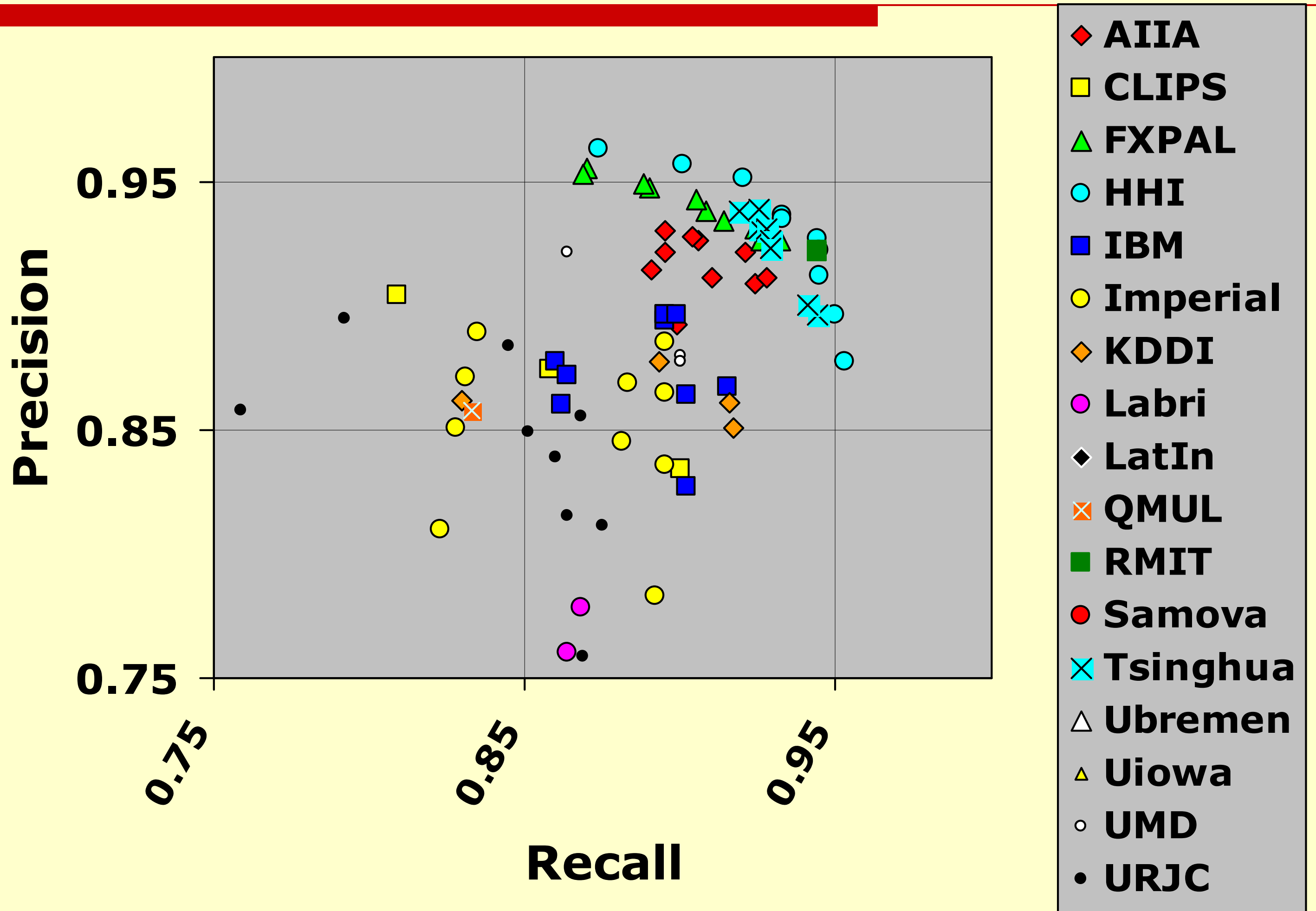


Precision
Recall
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.5
0.6
0.7
0.8
0.9
1
CLIPS
Fudan
FIU
FXPal
HKPU
IBM
IITB
Imperial
KDDI
LaBri
Marburg
Motorola
NICTA
RMIT
Tsinghua
TUDelft
UIowa
Umodena
URJC
USP
CityU-HK

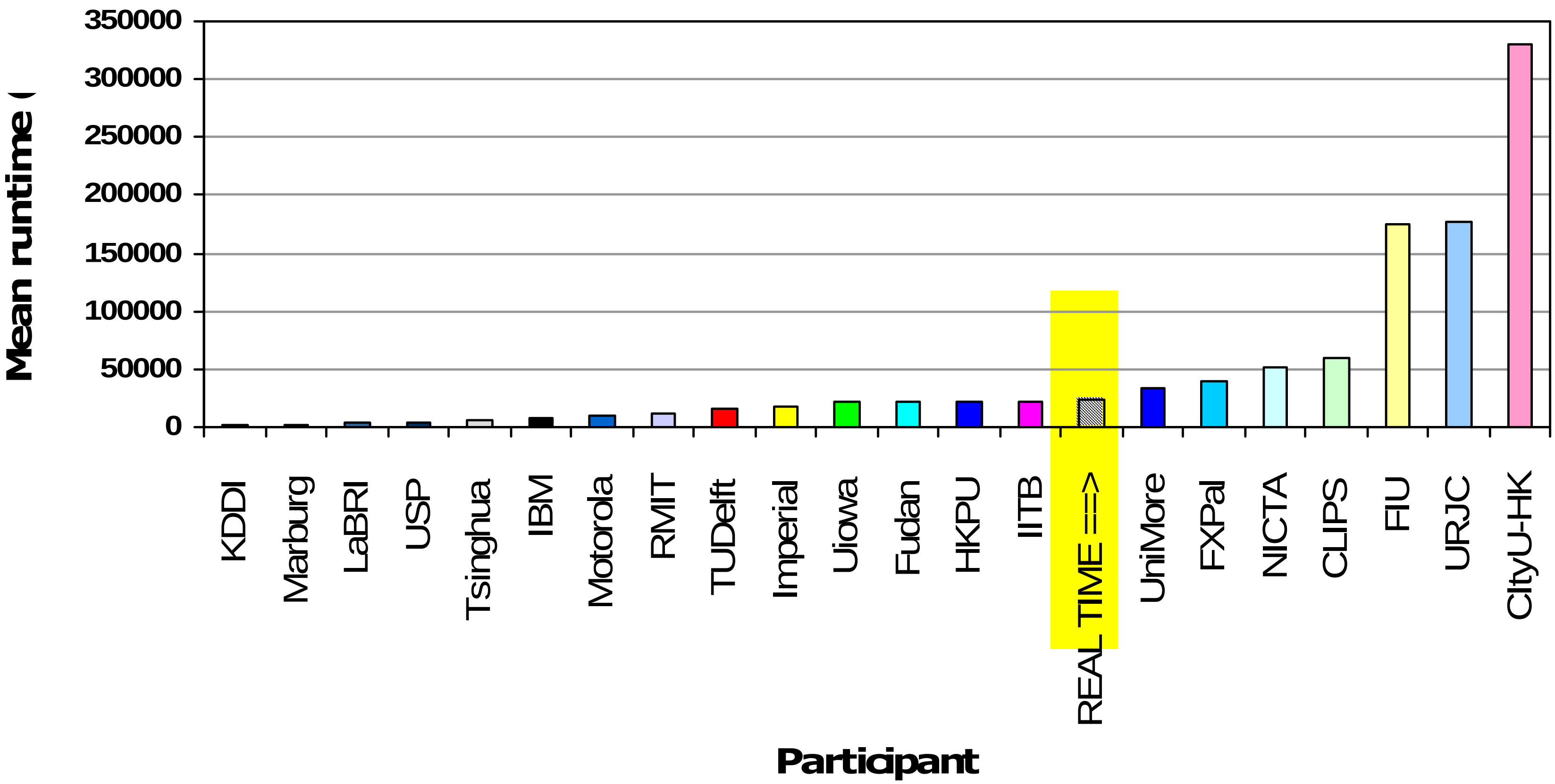
Cuts (zoomed again)



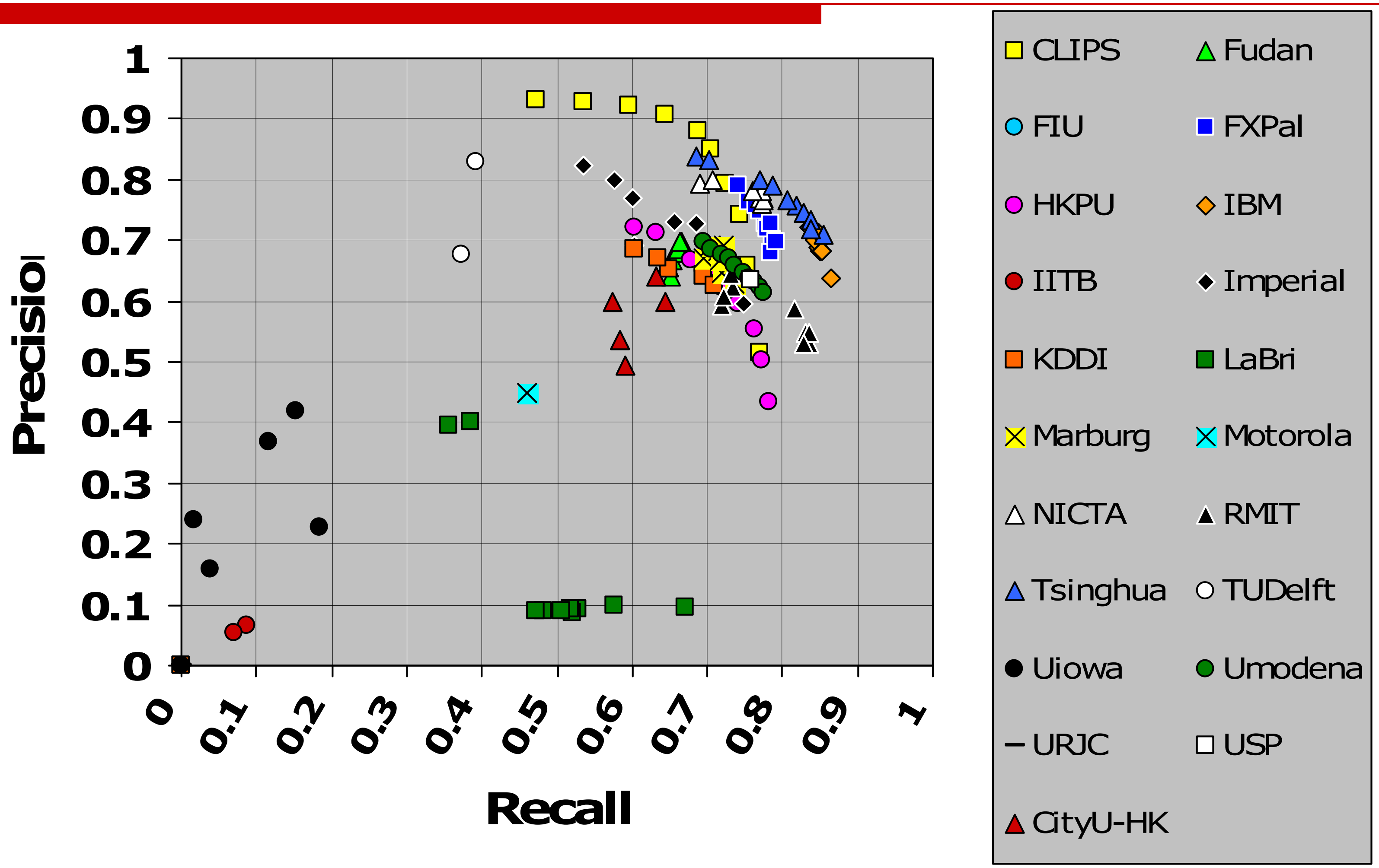
2004: Cuts (zoomed again)



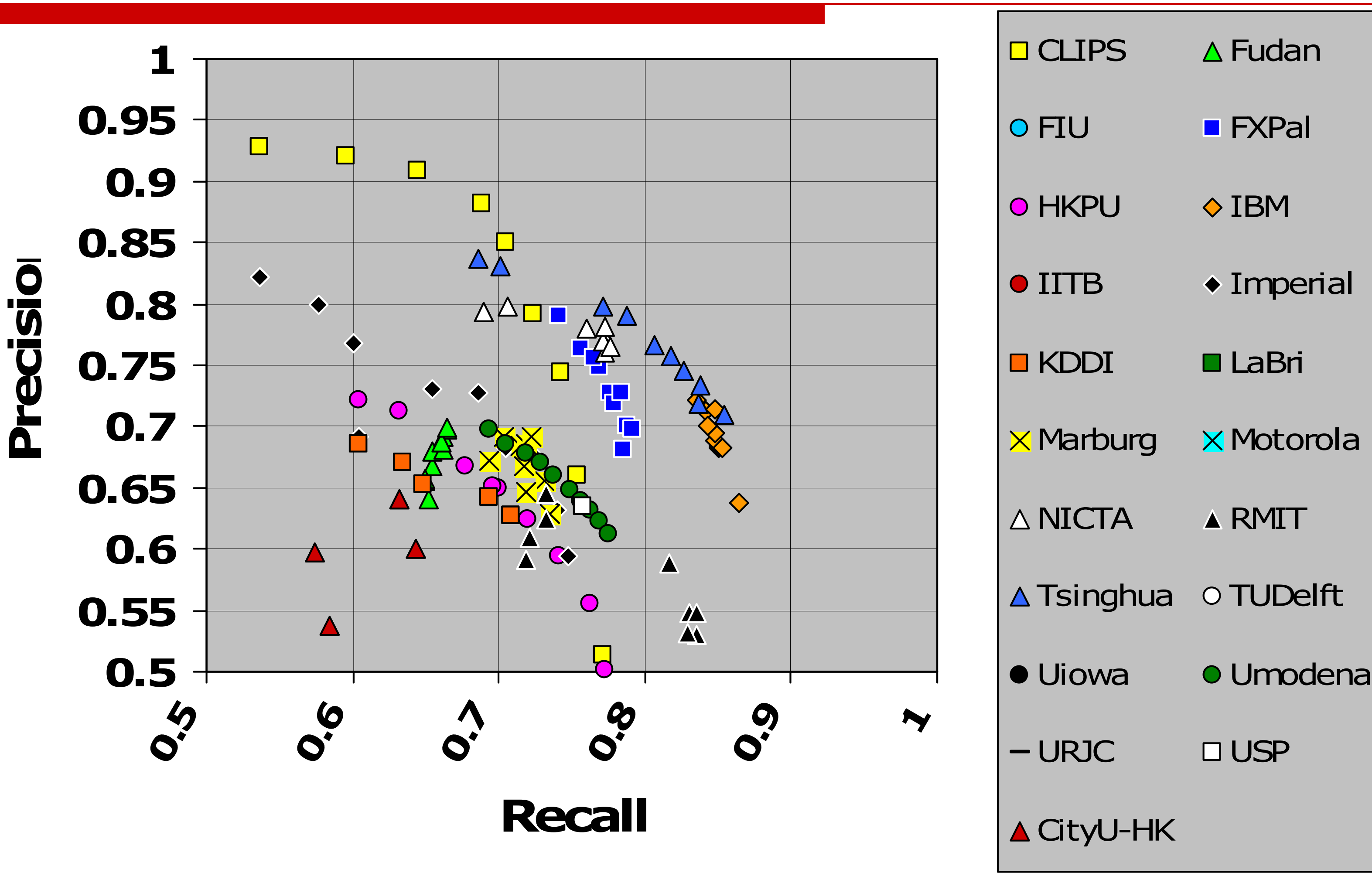
Mean runtime in seconds



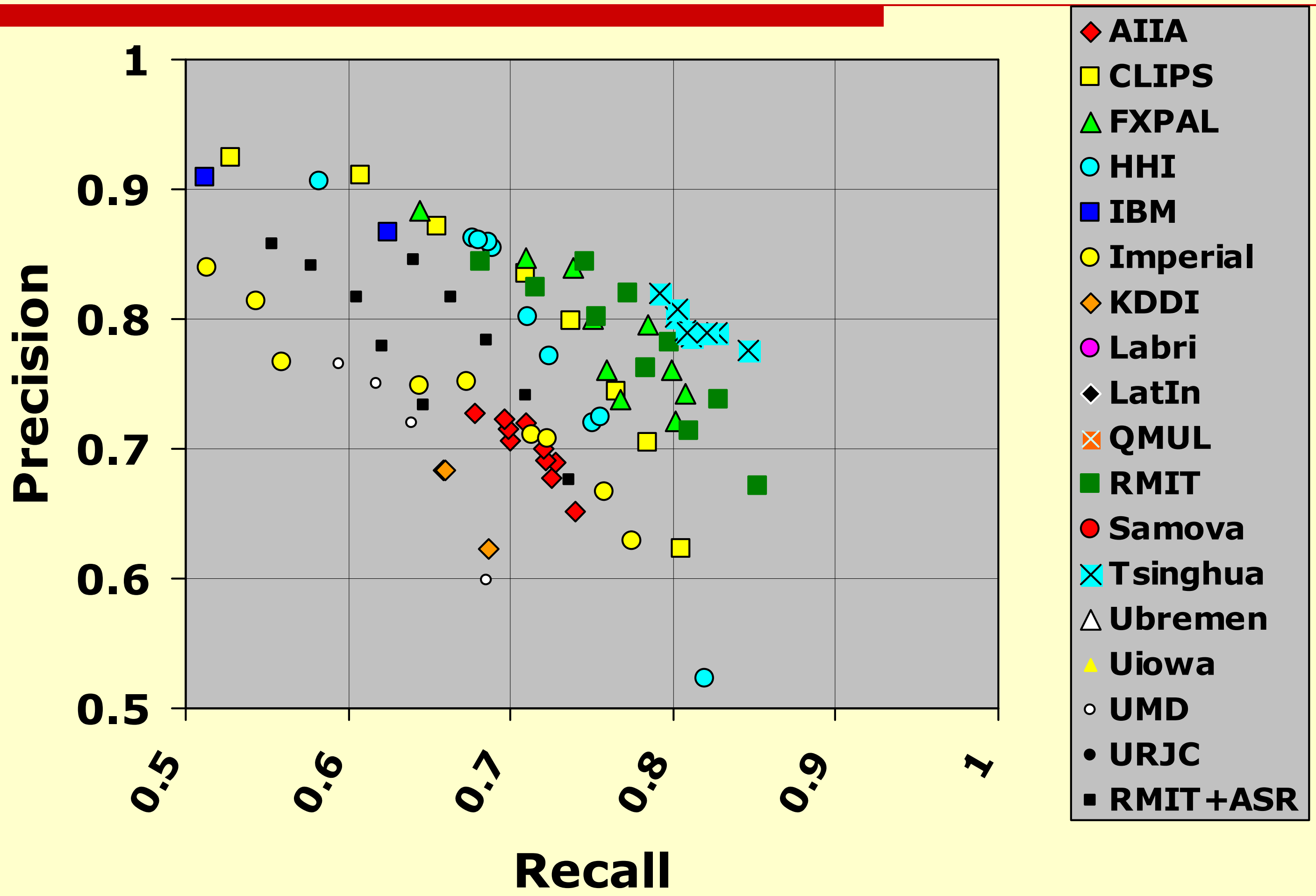
Gradual transitions



Gradual transitions (zoomed)



2004: Gradual transitions (zoomed)



Evaluation Measures

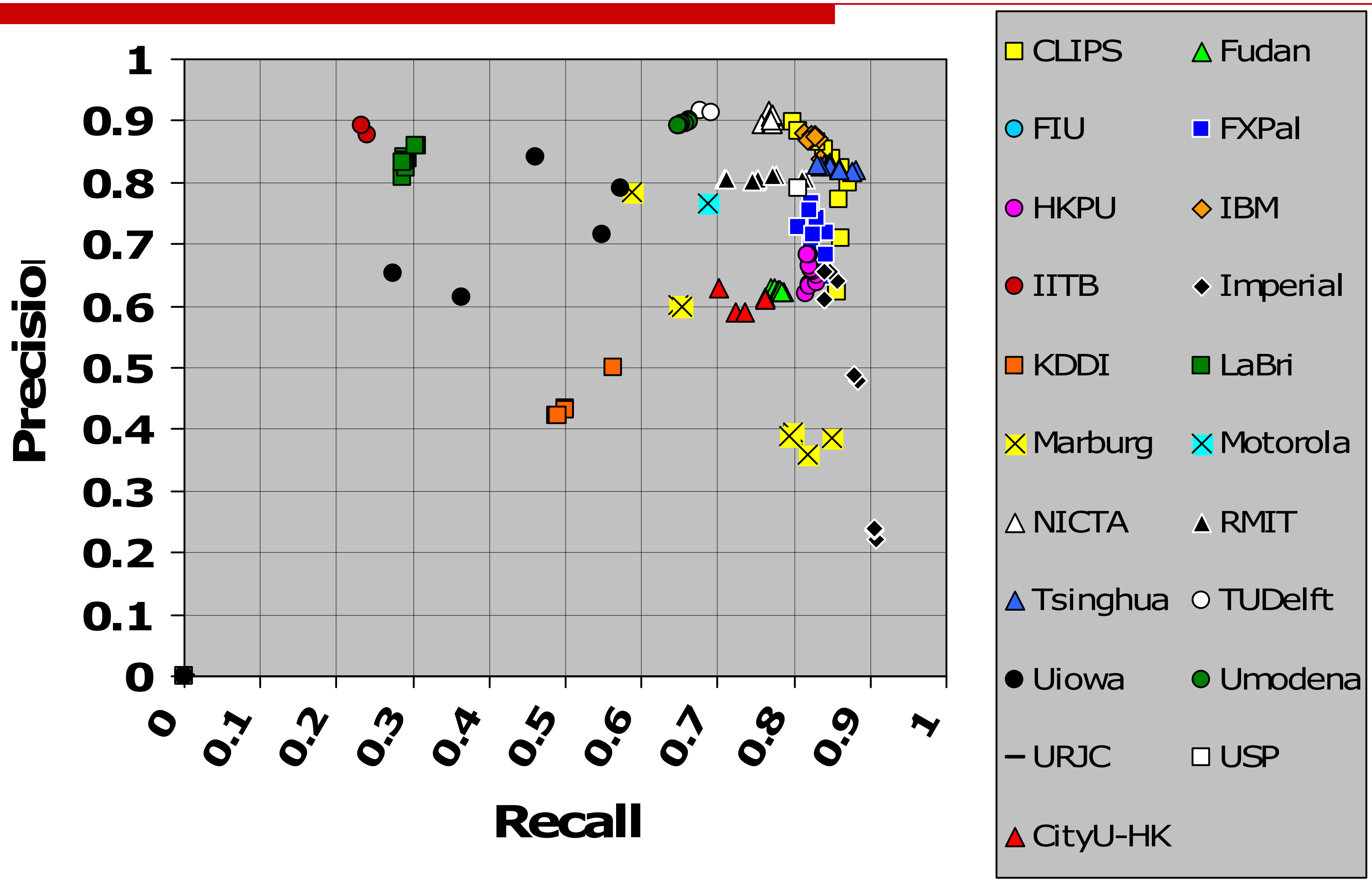
$$\text{Precision} = \frac{\text{\# Transitions Correctly Reported}}{\text{\# Transitions Reported}}$$

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

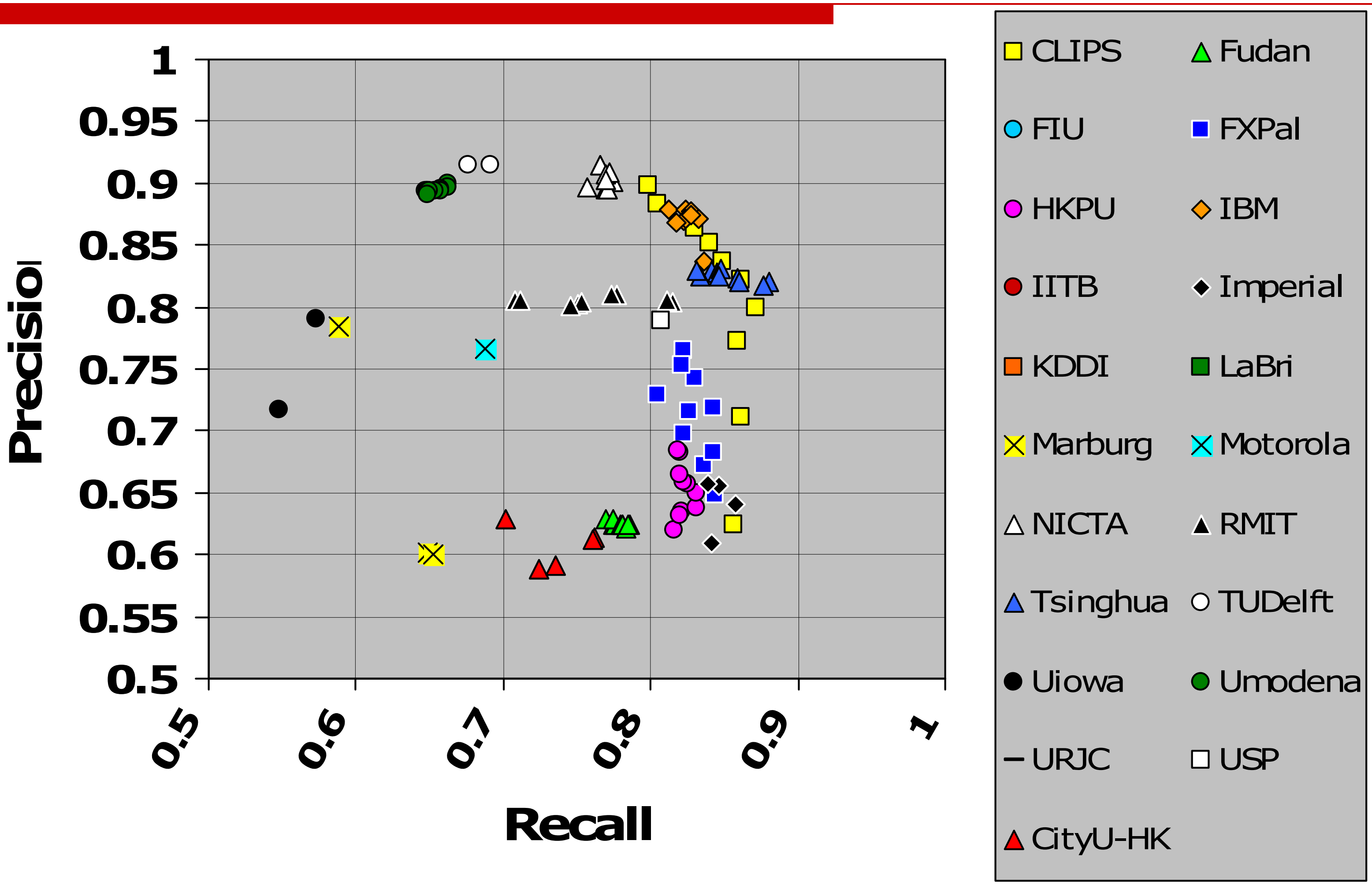
$$\text{Frame Precision} = \frac{\text{\# Frames Correctly Reported in Detected Transitions}}{\text{\# Frames reported in Detected Transitions}}$$

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

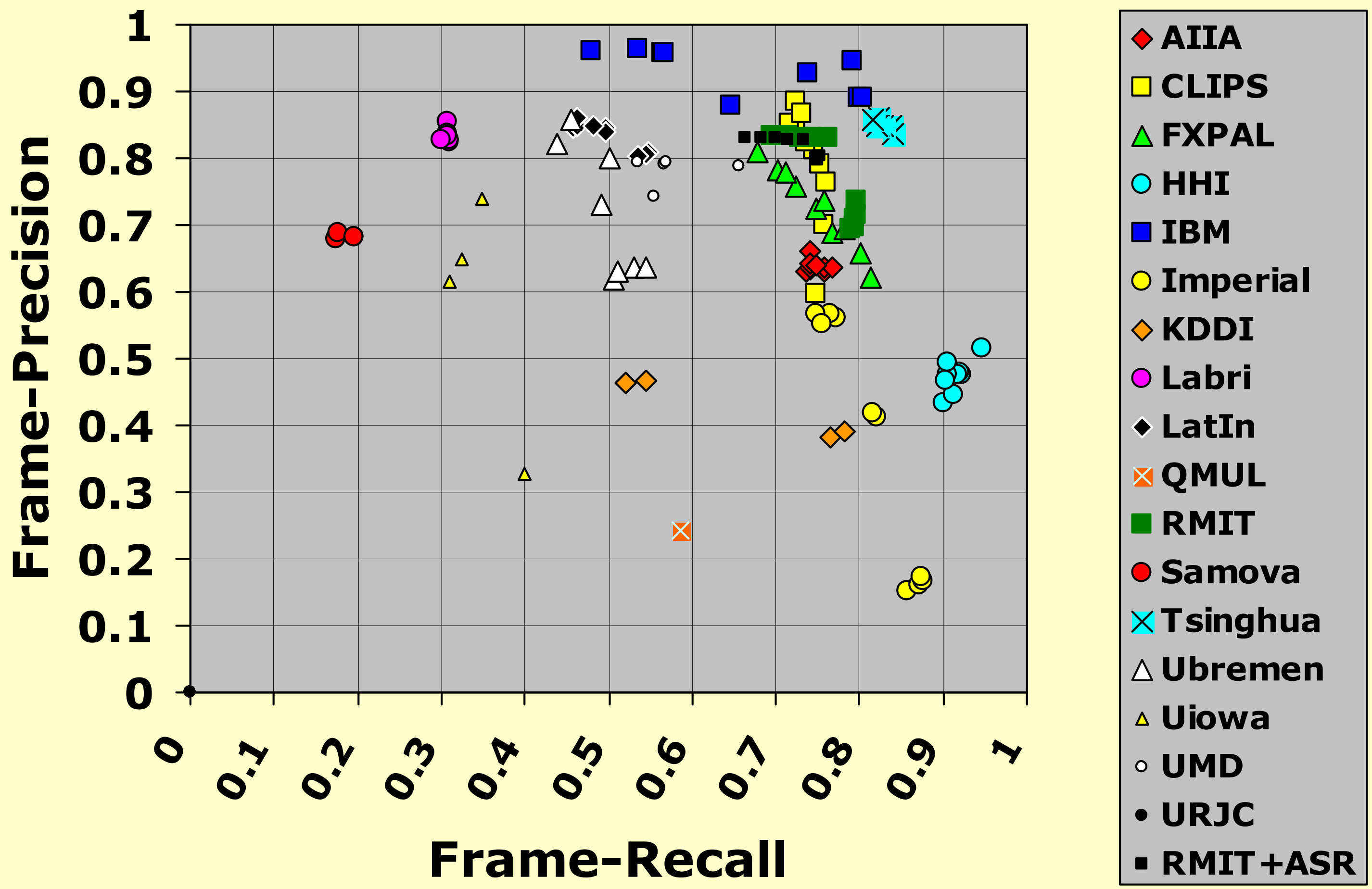
Gradual transitions (Frame-P & R)



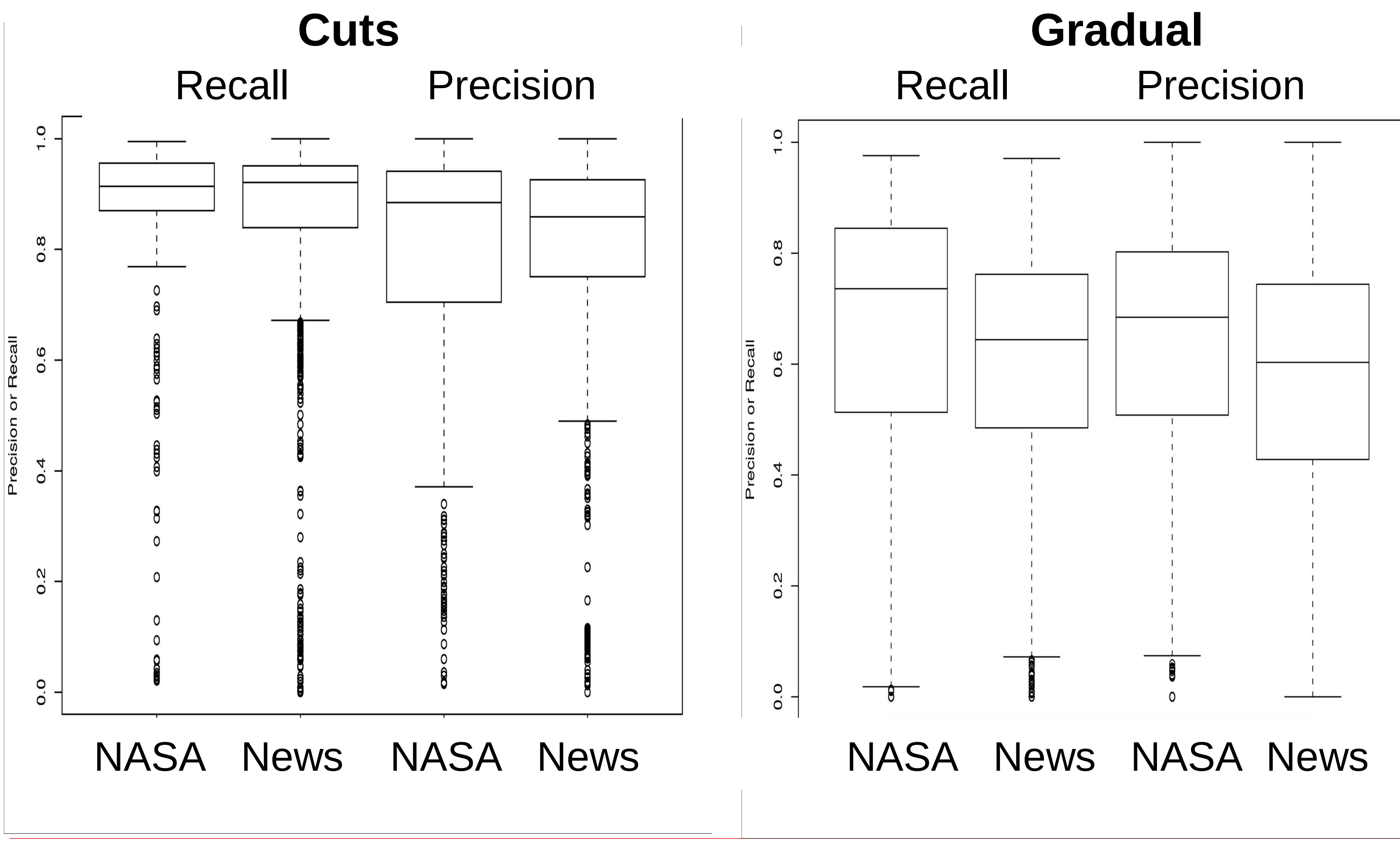
Gradual transitions: Frame-P & R (zoomed)



2004: Gradual Transitions (Frame-P&R)



Results for News versus NASA videos – distribution of per-file recall and precision by source type



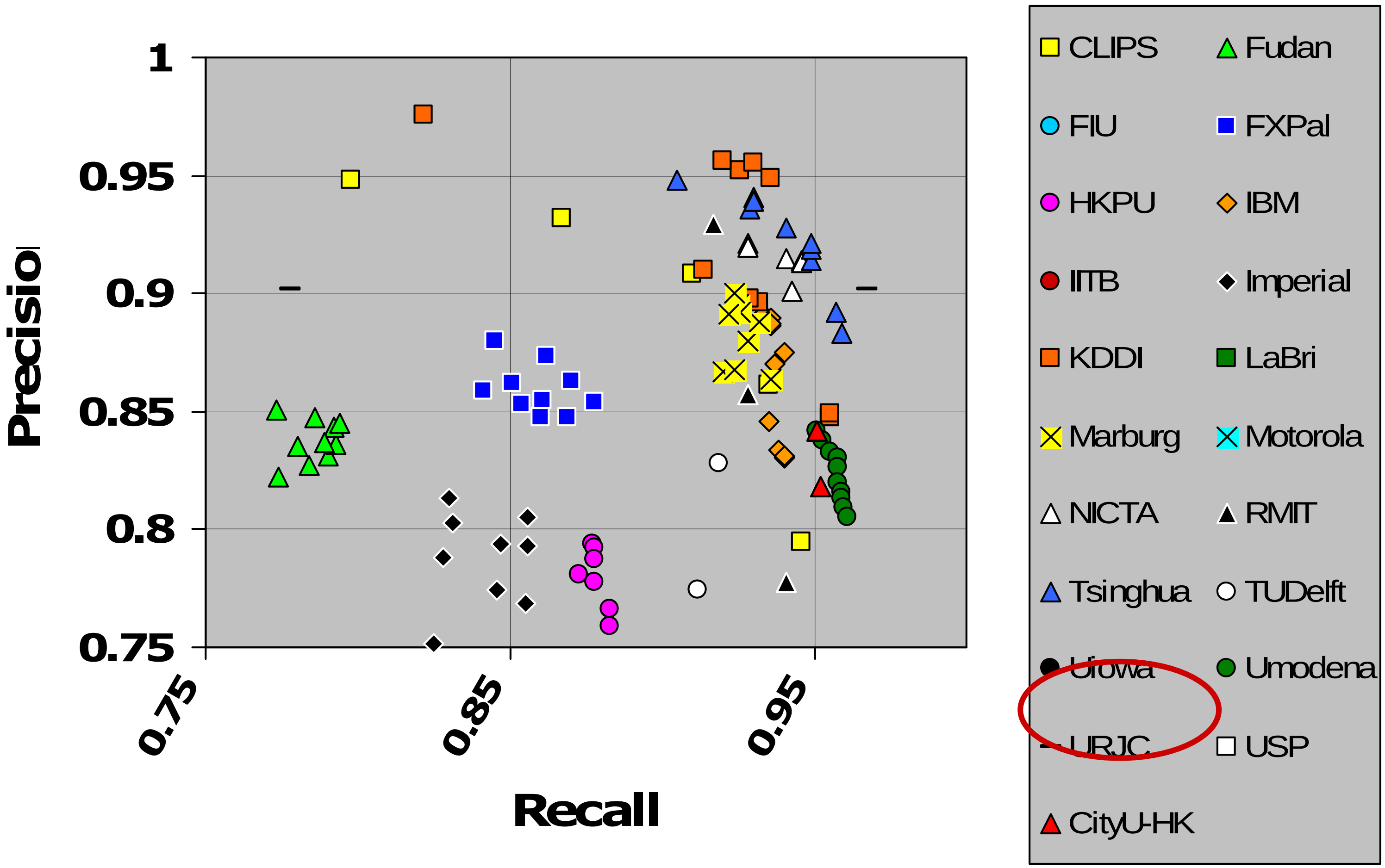
Approaches

A roller-coaster through 21 groups’ submitted runs;

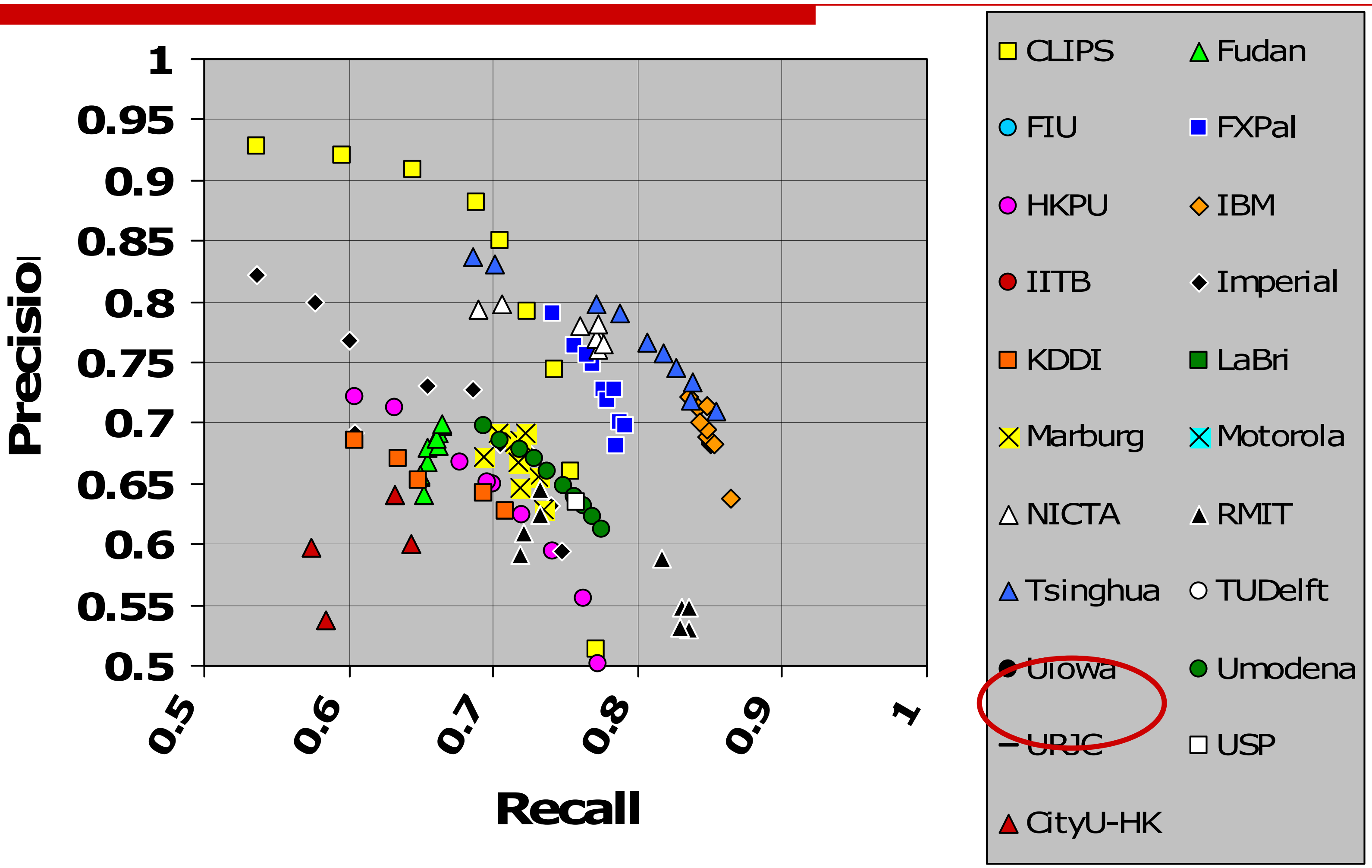
1. City University of Hong Kong

- Approach
  - Spatio-temporal (SD) slides are time vs. space representations of video - shot transition types (cuts, dissolves) appear in SDs with certain characteristics; Gabor features for motion texture and SVM for binary classification;
- Features
  - Expends previous (ACM MM) approach by including flash detection and extra visual features to discriminate GTs
- Performance
  - Because of image processing and SVM it is expensive;
- Results

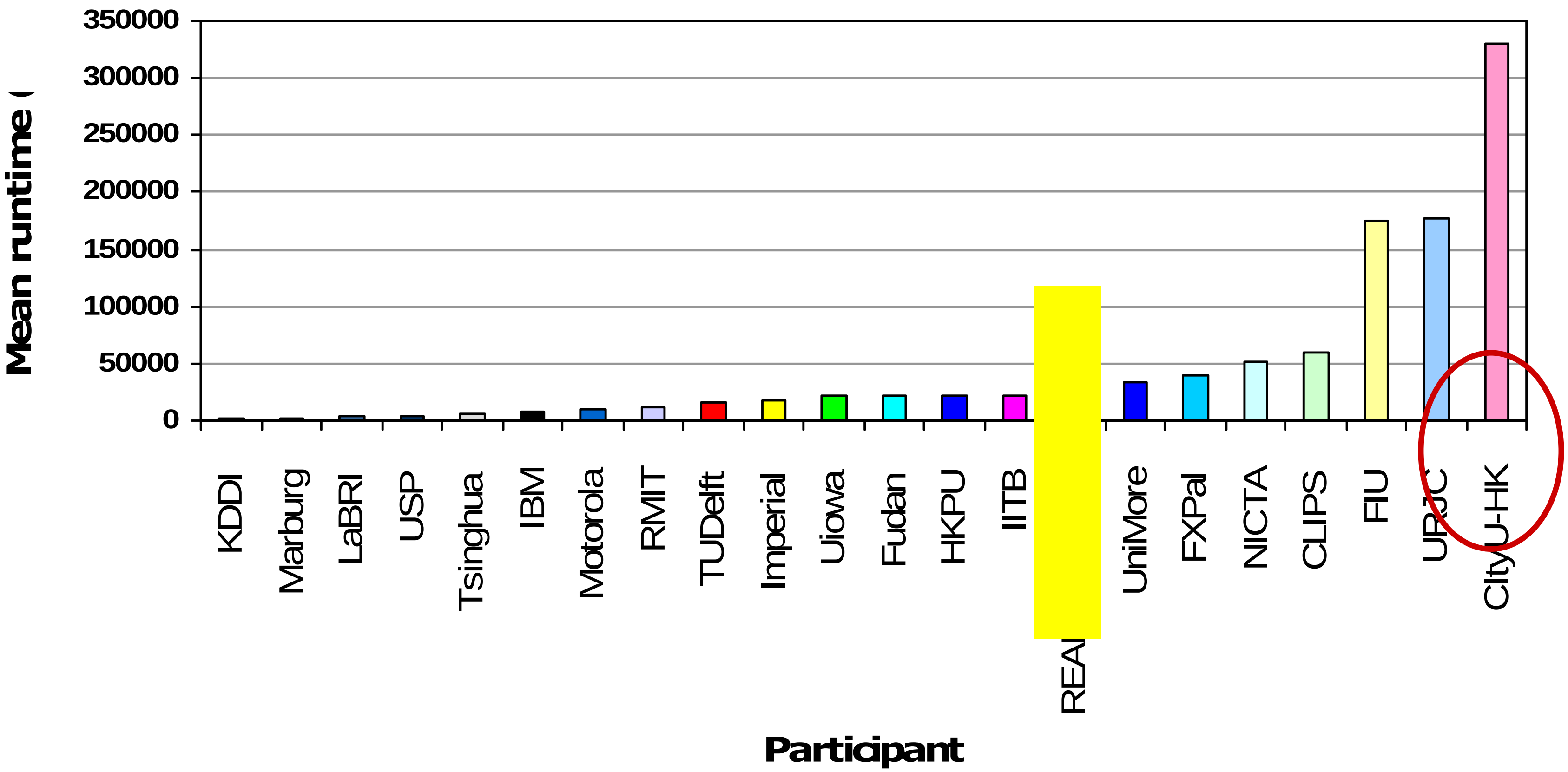
Cuts (zoomed again)



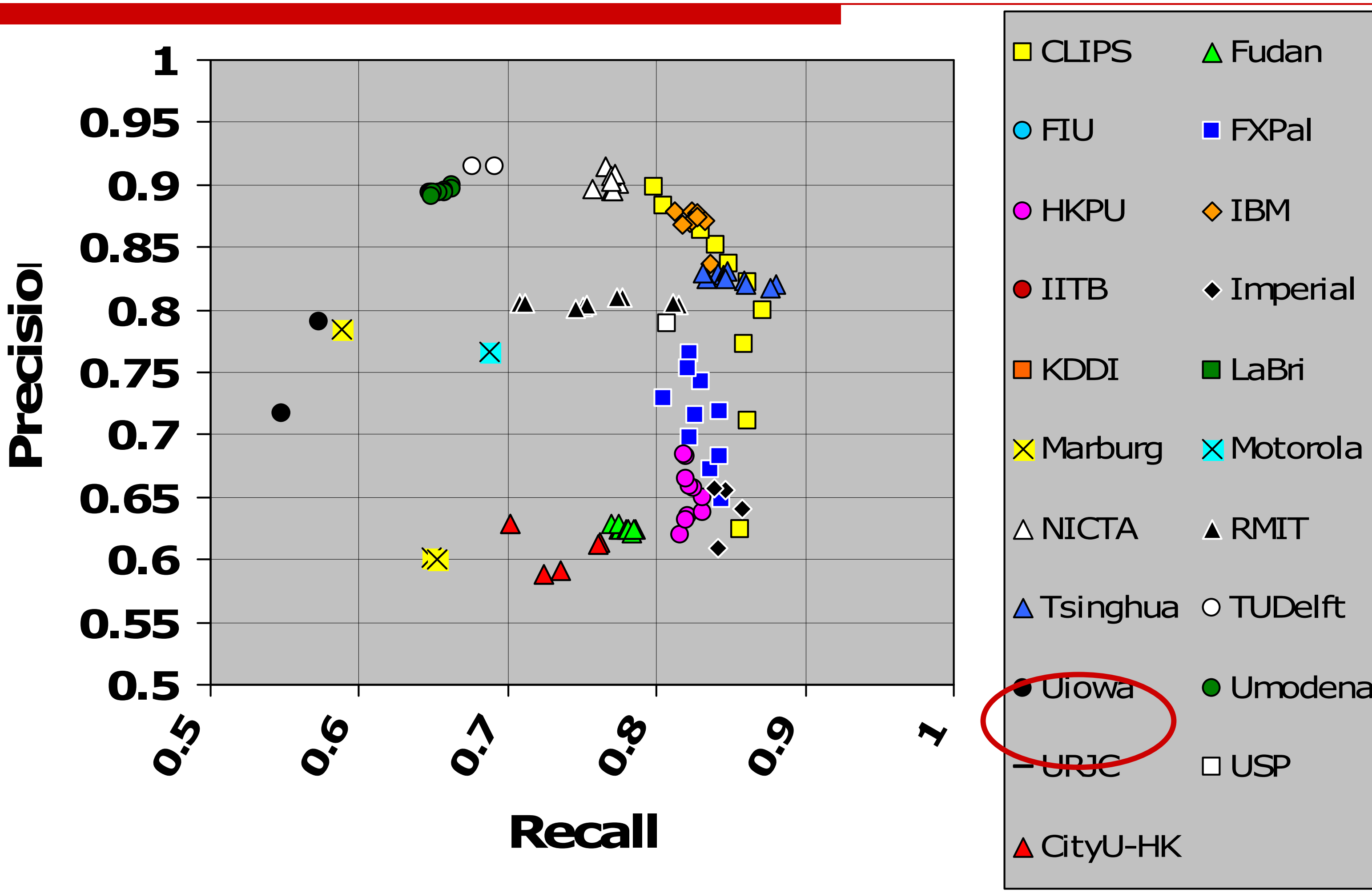
Gradual transitions (zoomed)



Mean runtime in seconds



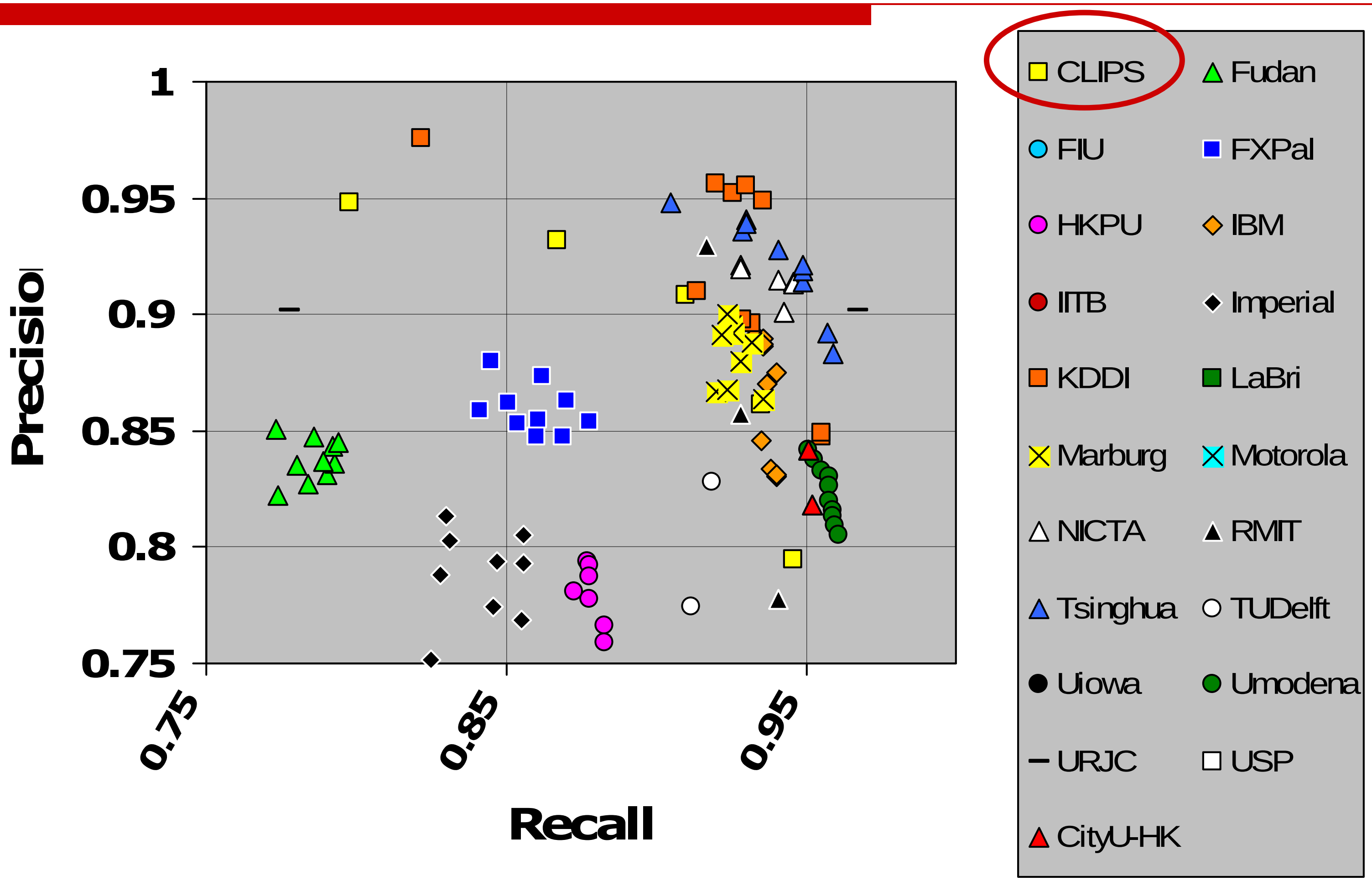
Gradual transitions: Frame-P & R (zoomed)



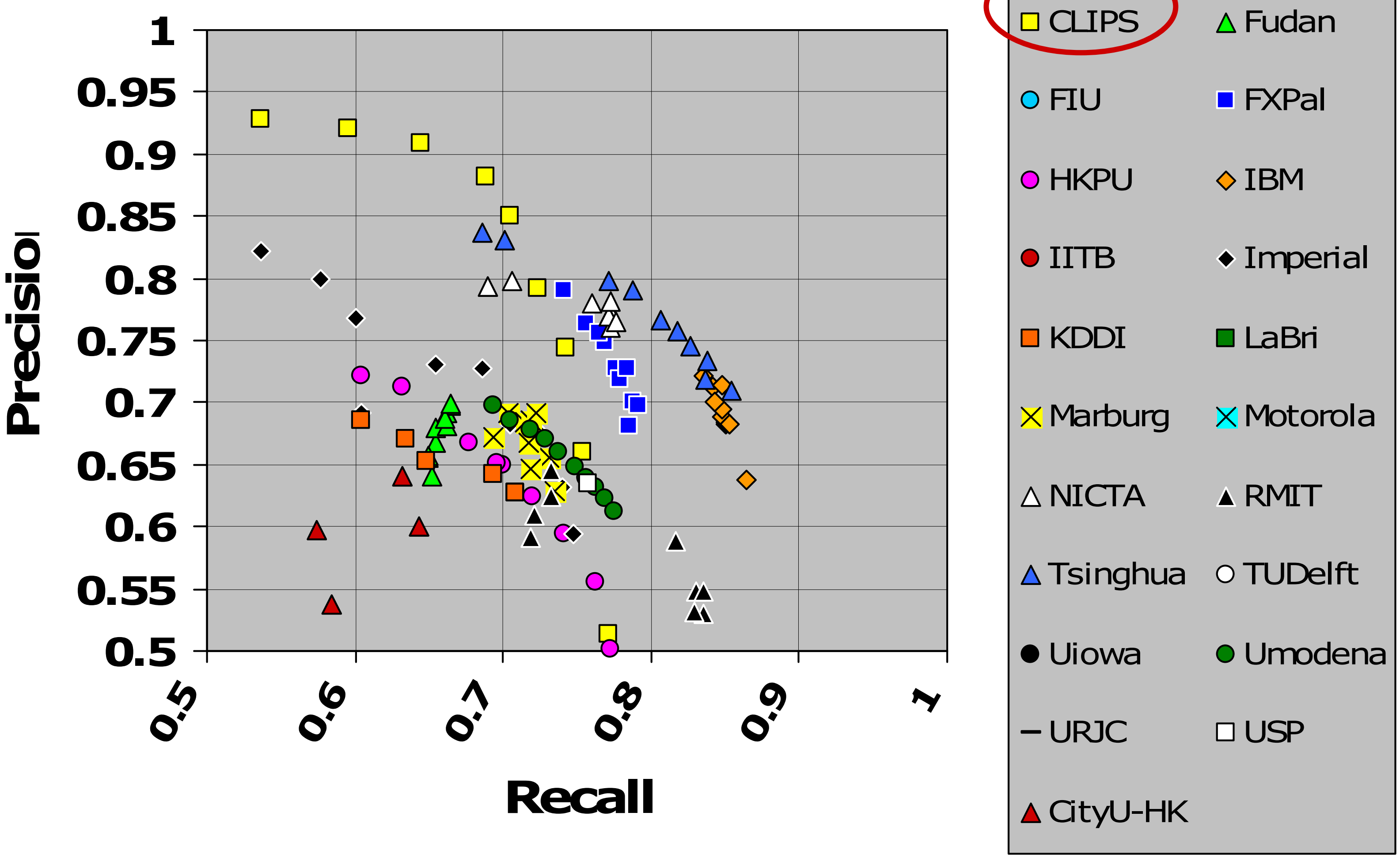
2. CLIPS-IMAG, LSR-IMAG, Laboratoire LIS

- Approach
  - Appears to be a re-run of 2004 system, which was a re-run of 2003 (thanks for doing this) - emphasis was on features.
- Features
  - Detect cuts by image comparisons after motion compensation and GTs by comparing norms of first and second temporal derivatives of the images;
- Performance
  - About real-time, good on GTs;
- Results

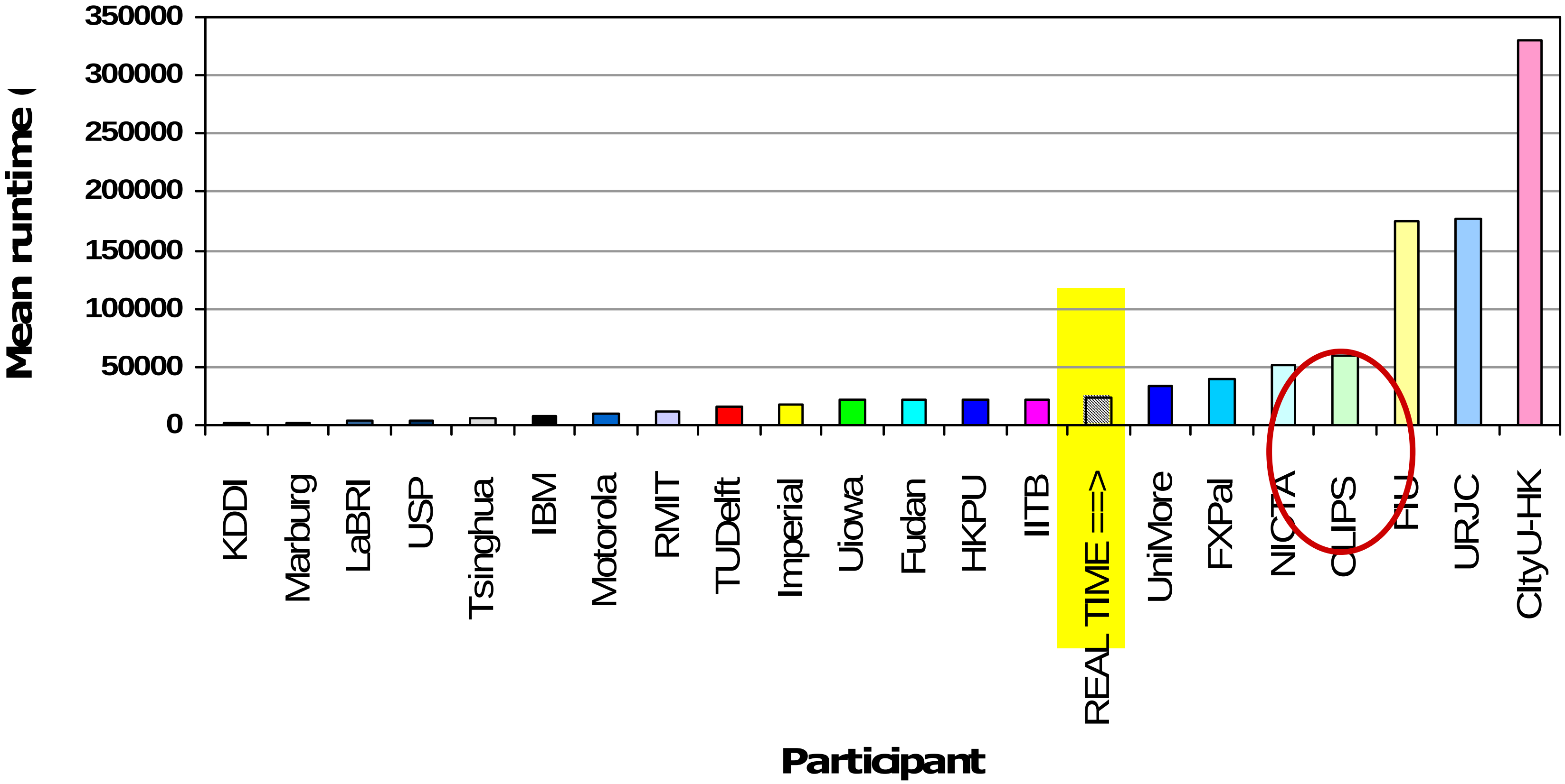
Cuts (zoomed again)



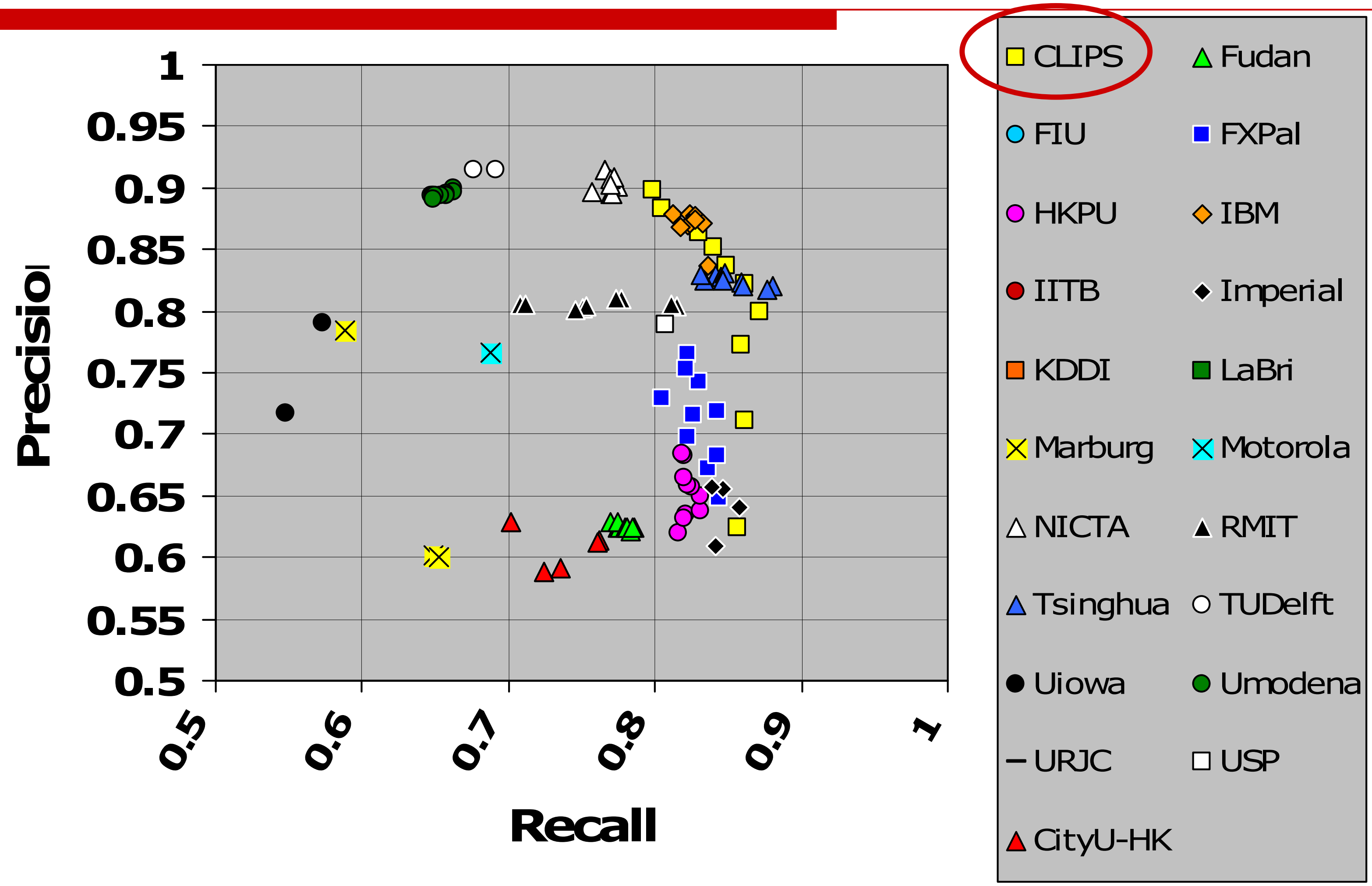
Gradual transitions (zoomed)



Mean runtime in seconds



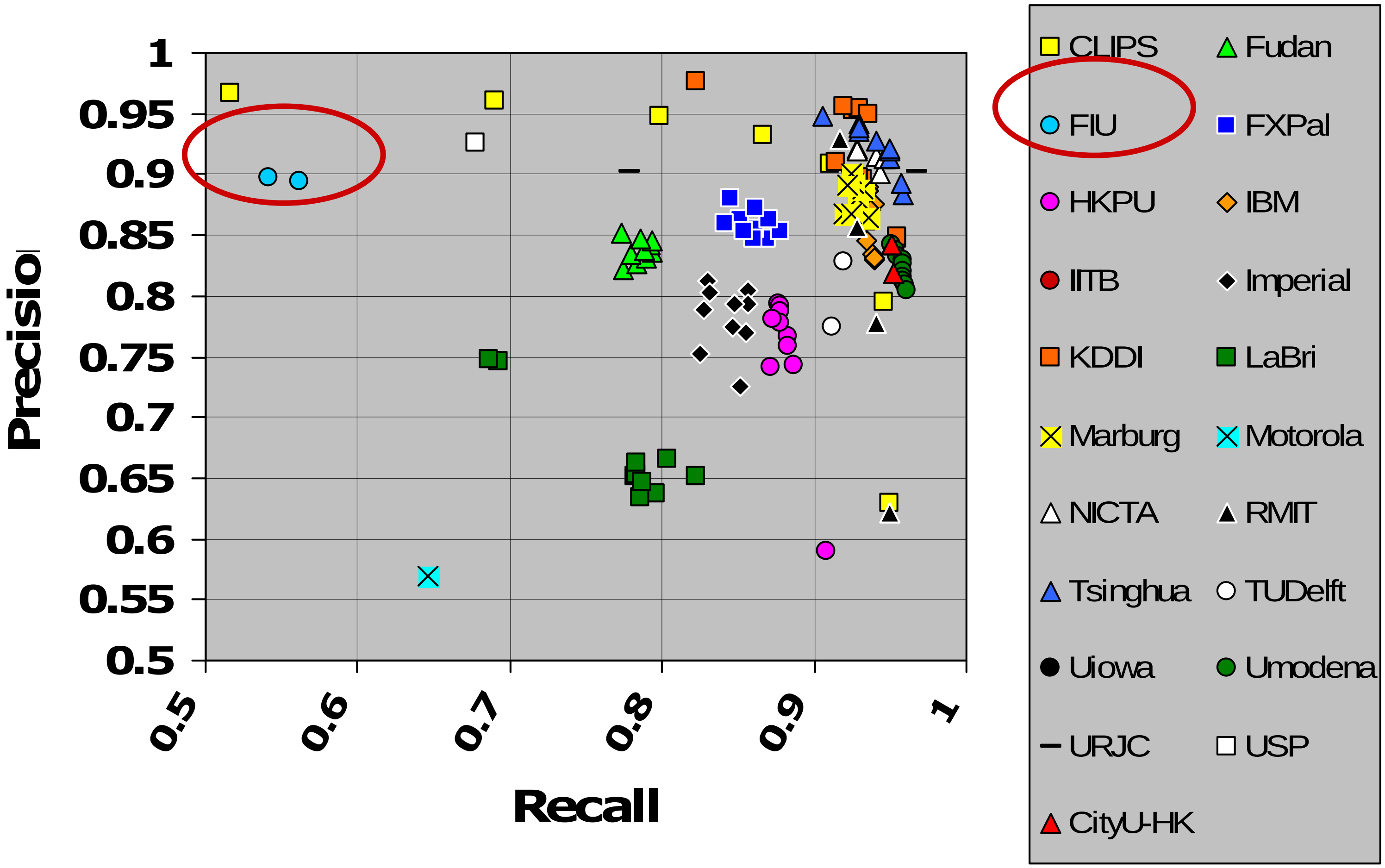
Gradual transitions: Frame-P & R (zoomed)



3. Florida International University

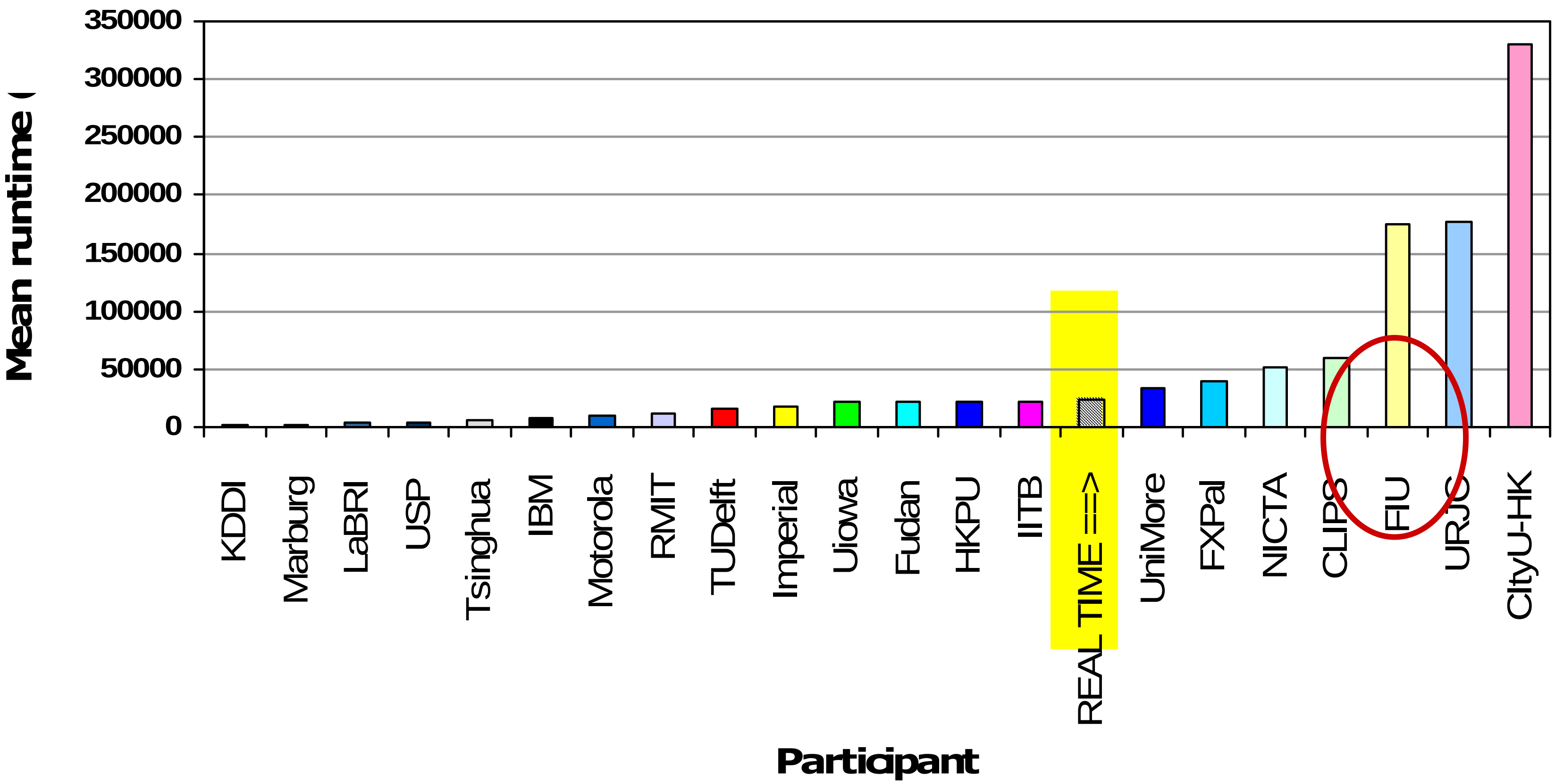
- Approach
  - Didn't submit a paper so we don't know !

Cuts (zoomed)



Precisio
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.5
0.6
0.7
0.8
0.9
1
Recall
CLIPS
Fudan
FIU
FXPal
HKPU
IBM
IITB
Imperial
KDDI
LaBri
Marburg
Motorola
NICTA
RMIT
Tsinghua
TUDelft
Uiowa
Umodena
URJC
USP
CityU-HK

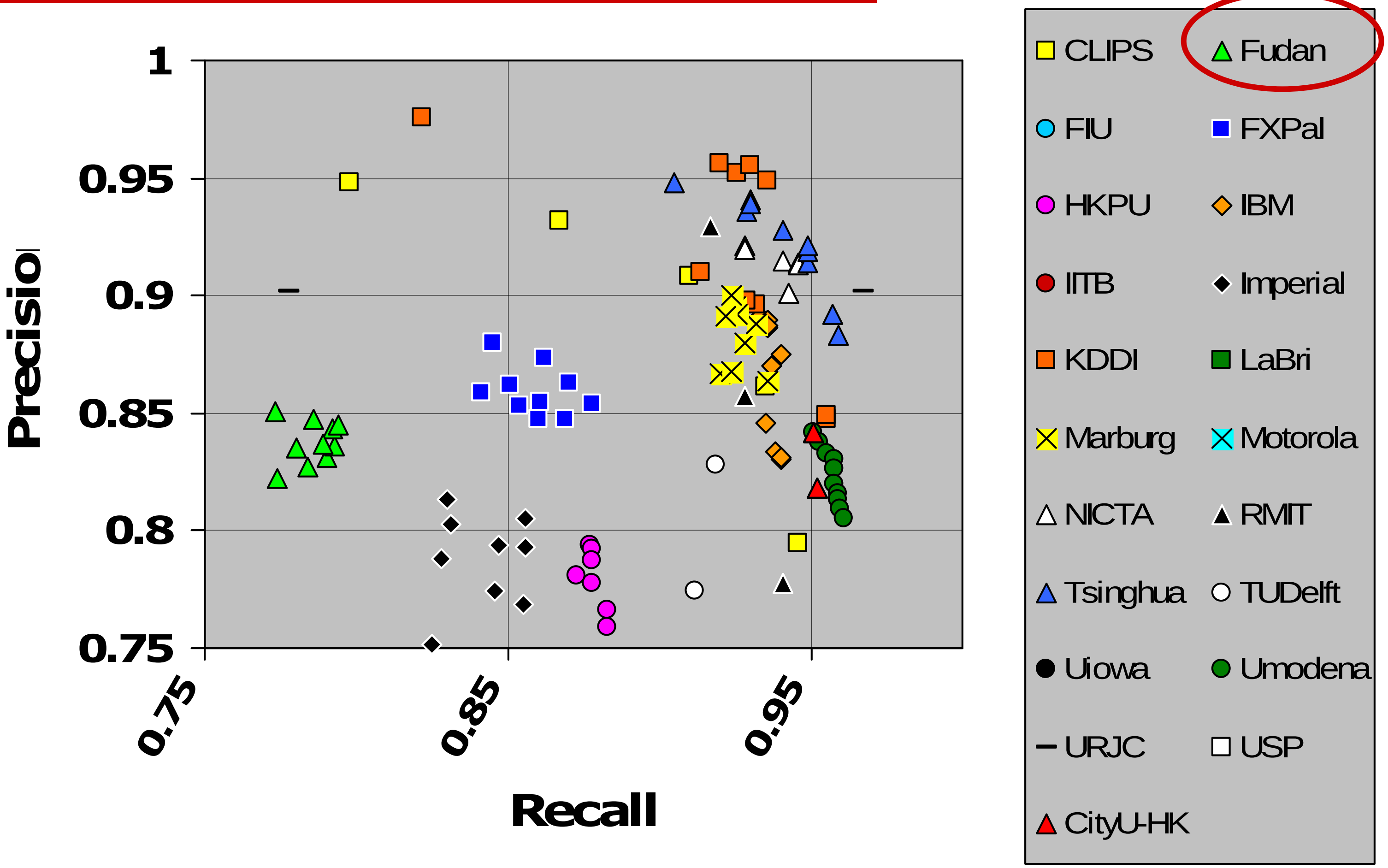
Mean runtime in seconds



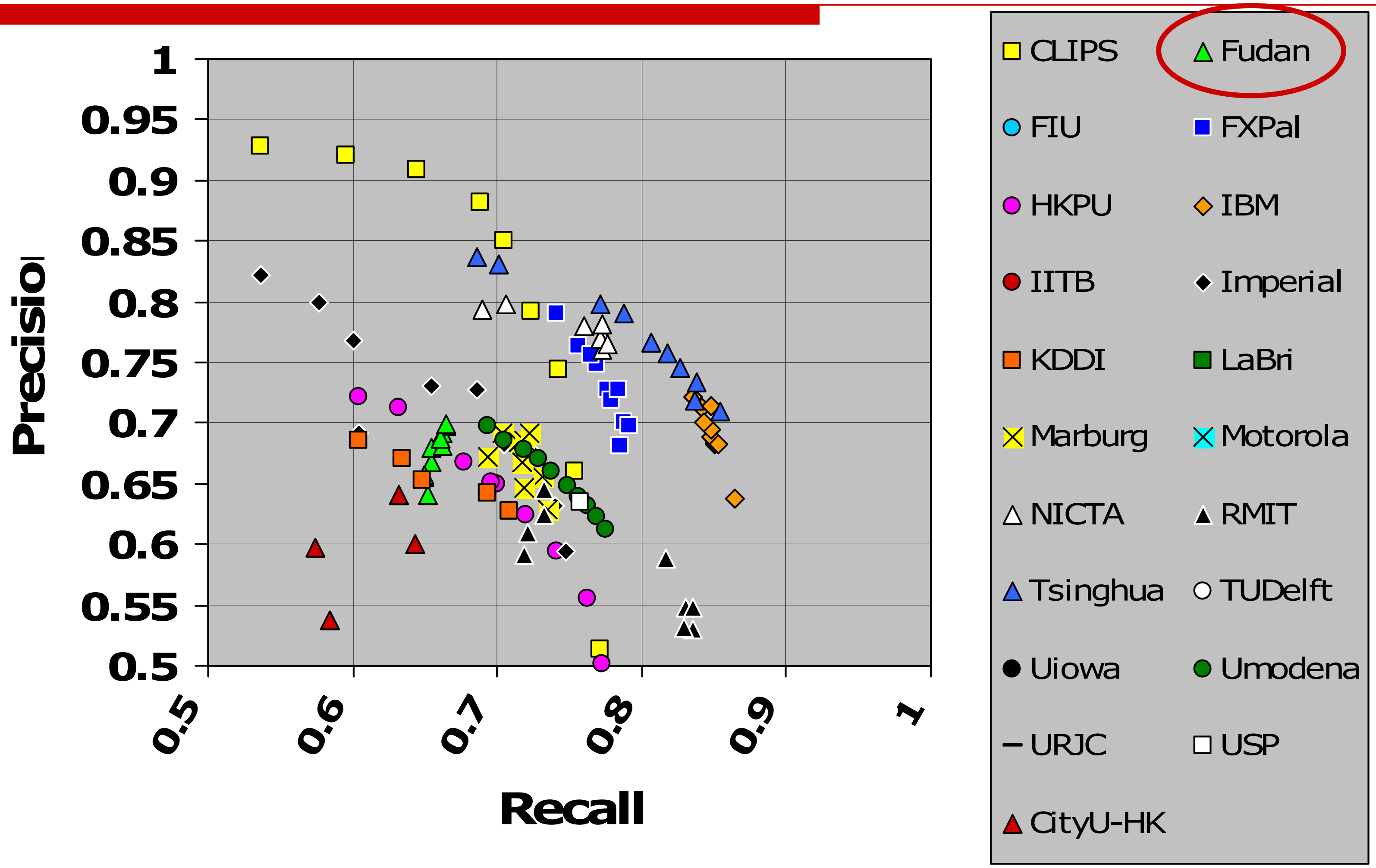
4. Fudan University

- Approach
  - Frame-frame similarities, vary thresholds, use SVM classifier;
  - Explore HSV vs. LAB colour spaces;
- Features
  - Fudan definition of a short GT is a cut, differs from TRECVid evaluation, hence results depressed;
- Performance
  - About mid-table in runtime and in accuracy;
- Results
  - No differences between colour spaces

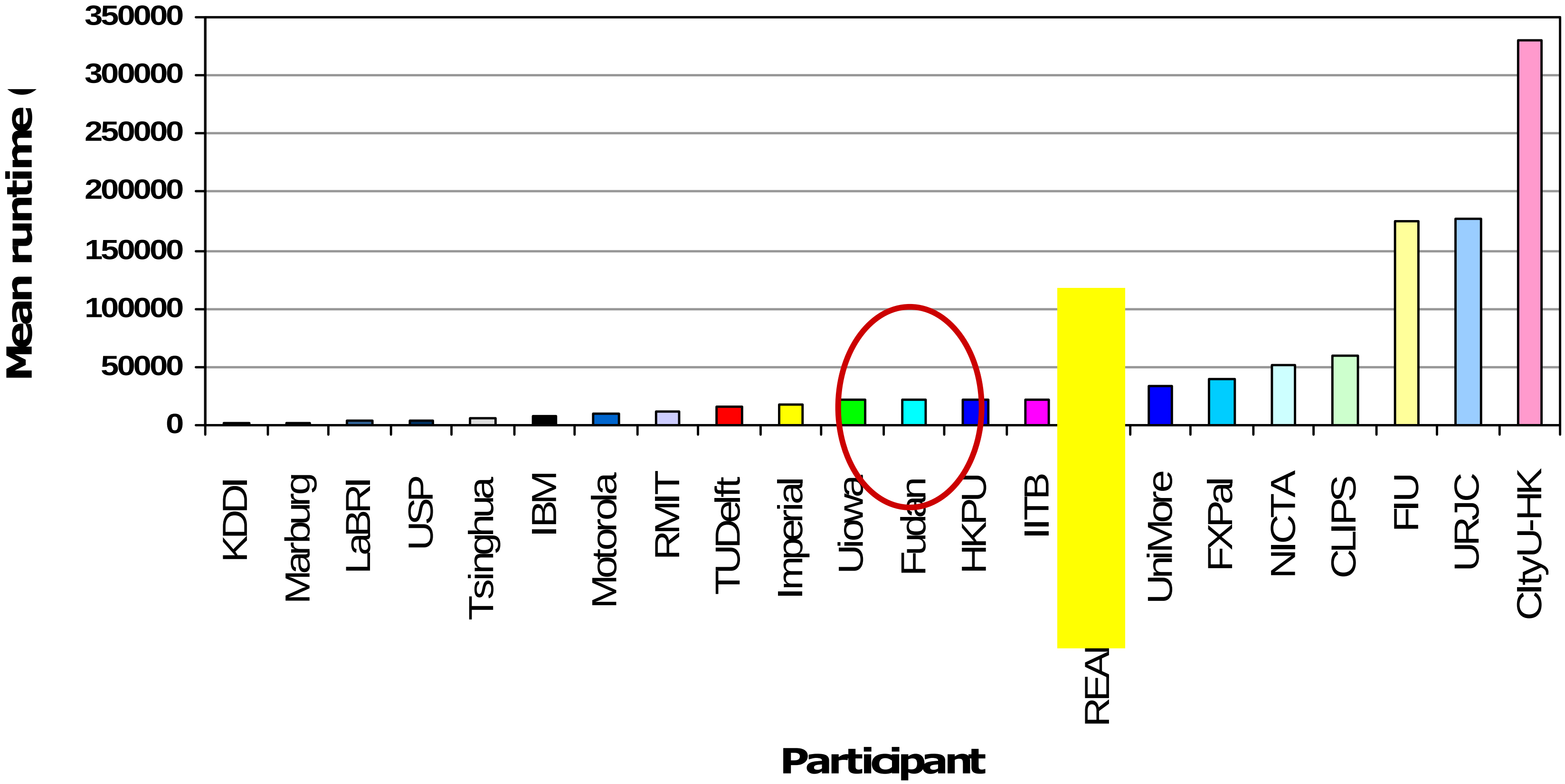
Cuts (zoomed again)



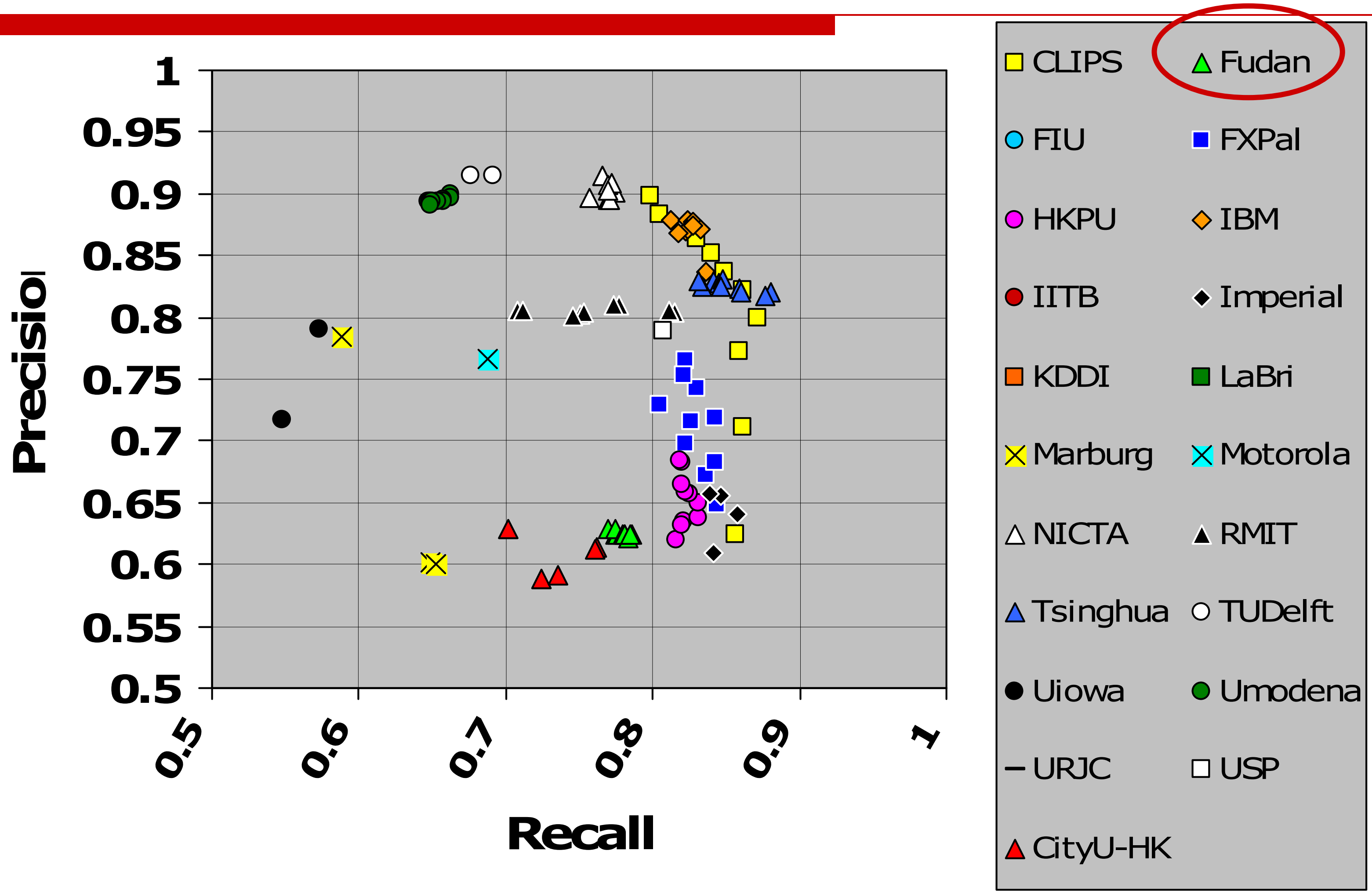
Gradual transitions (zoomed)



Mean runtime in seconds



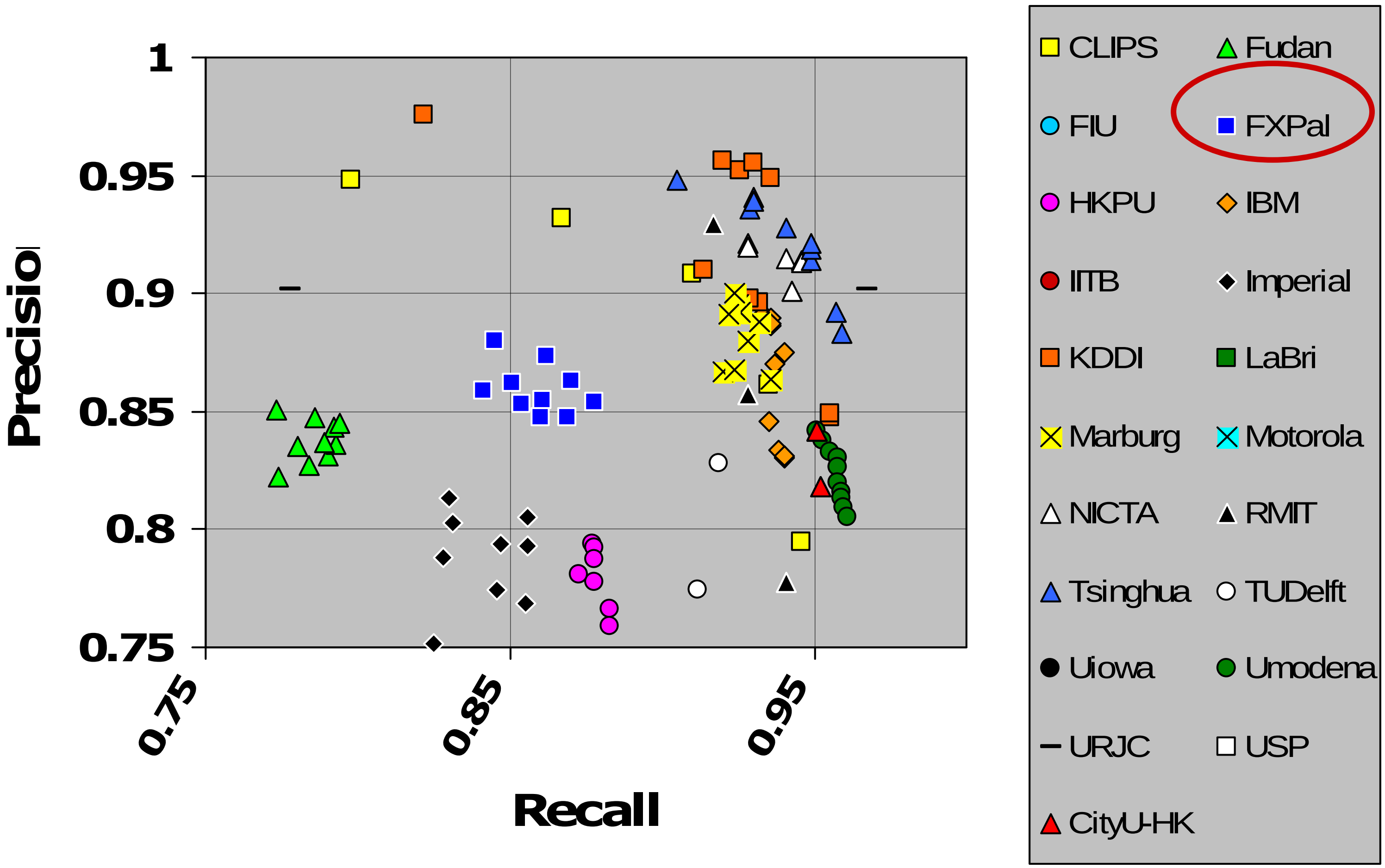
Gradual transitions: Frame-P & R (zoomed)



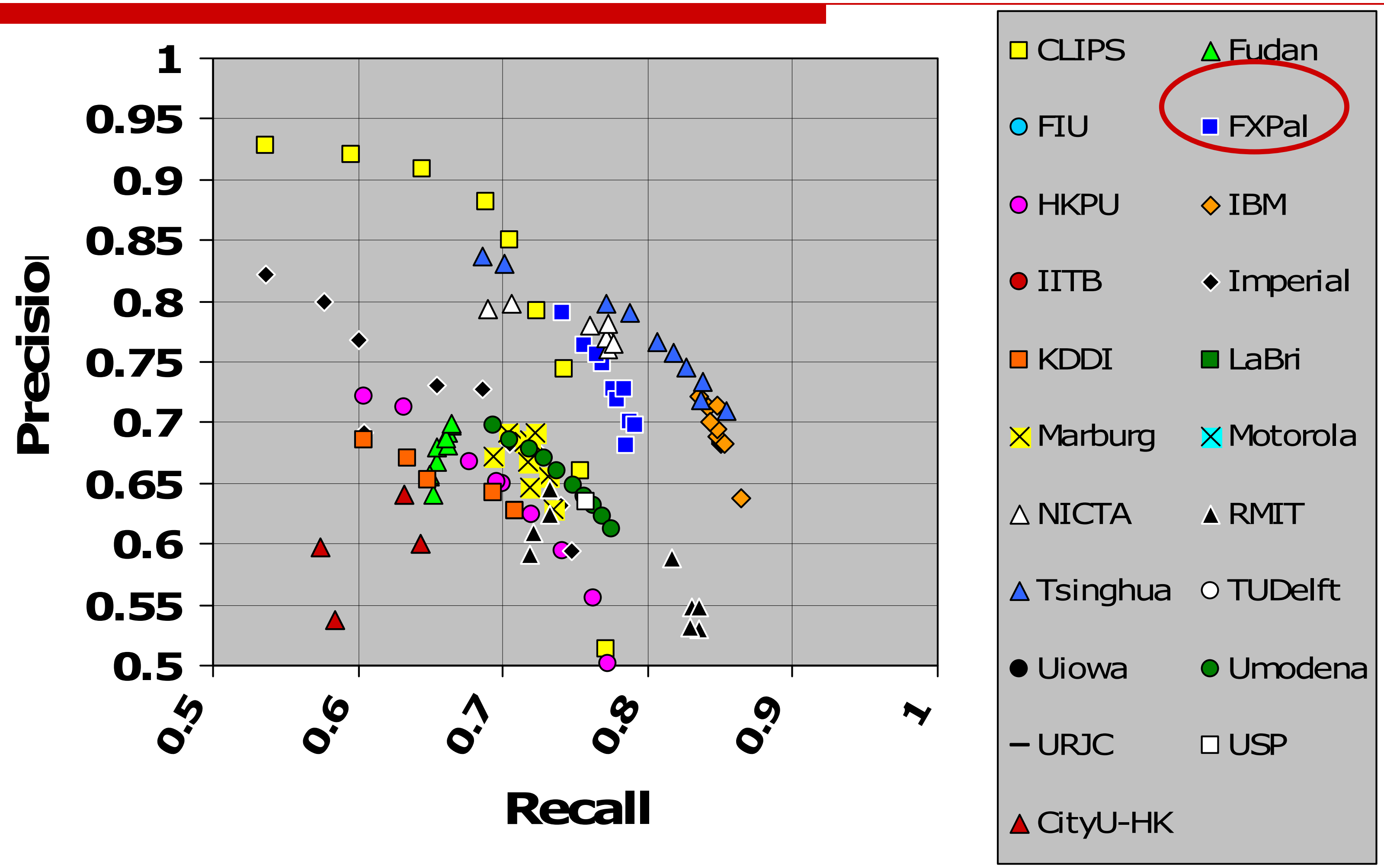
5. FX Palo Alto Laboratory

- Approach
  - Builds upon previous years with intermediate visual features derived from low-level image features for pairwise frame similarities over local and longer-distances;
  - Used as input to a kNN classifier;
  - Added information-theoretic secondary feature selection to select features used in classifier;
- Features
  - Feature selection/reduction yielded improved performances;
- Performance
  - Not as good as expected because sensitive to training data;
- Results

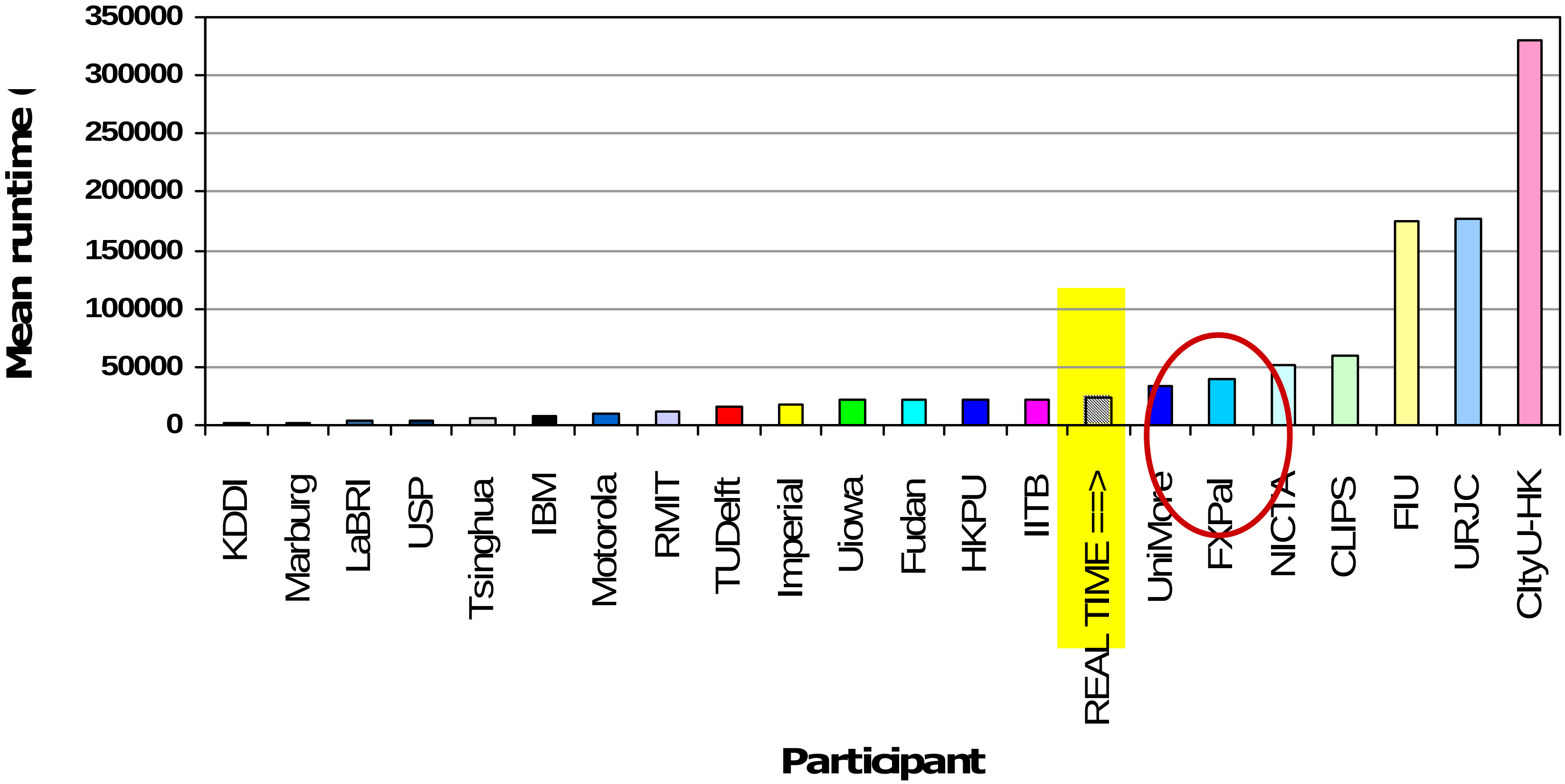
Cuts (zoomed again)



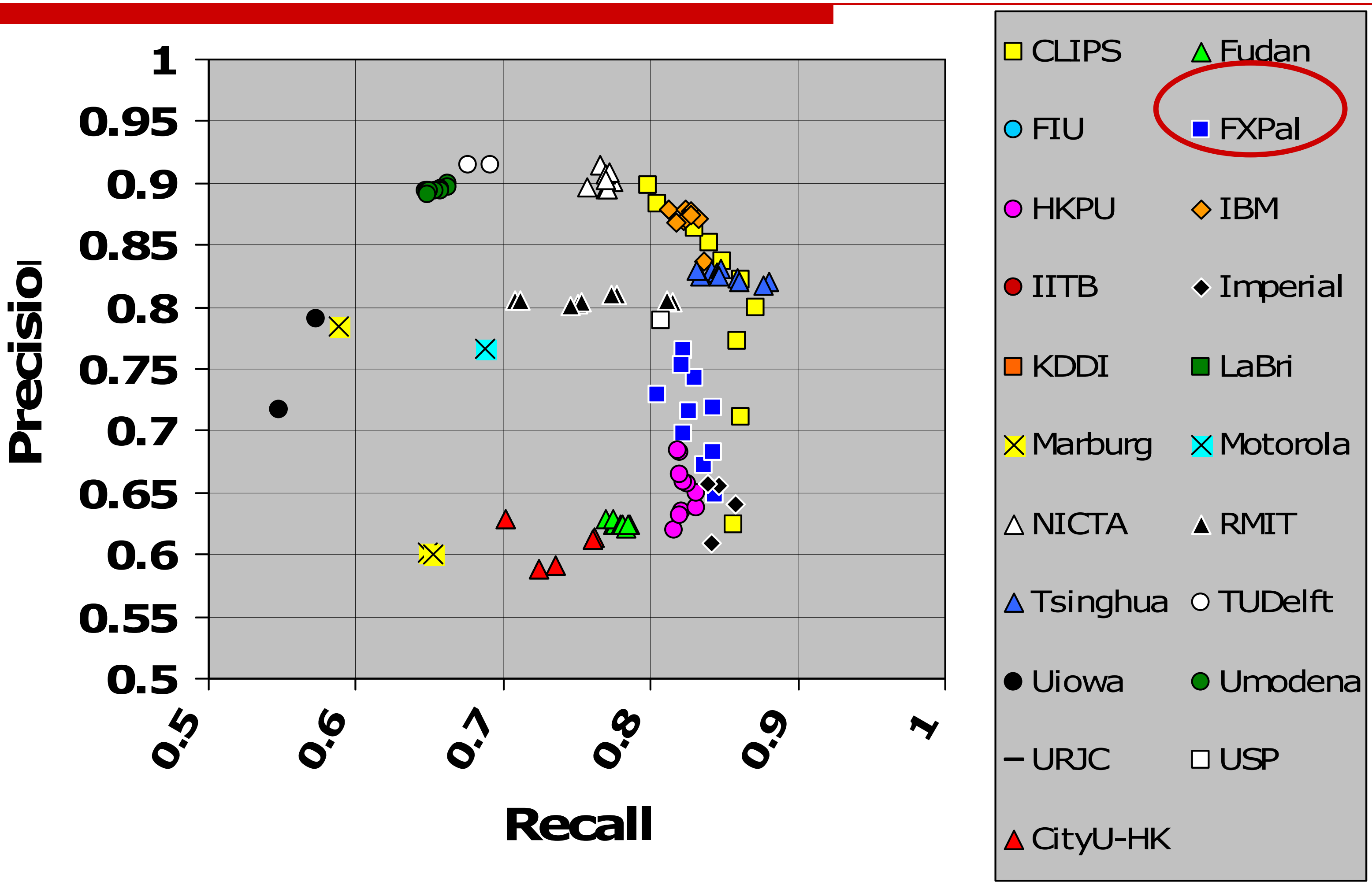
Gradual transitions (zoomed)



Mean runtime in seconds



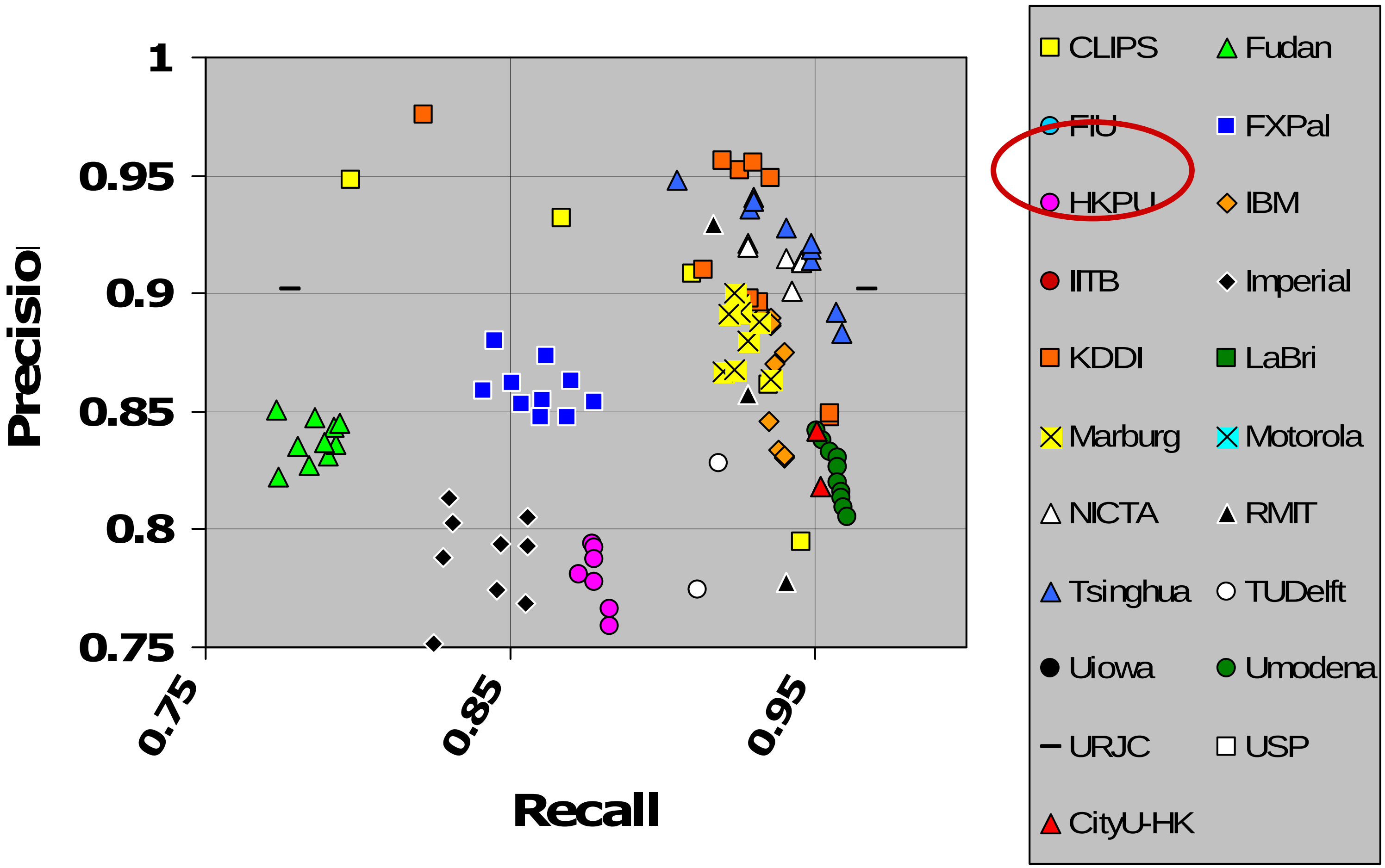
Gradual transitions: Frame-P & R (zoomed)



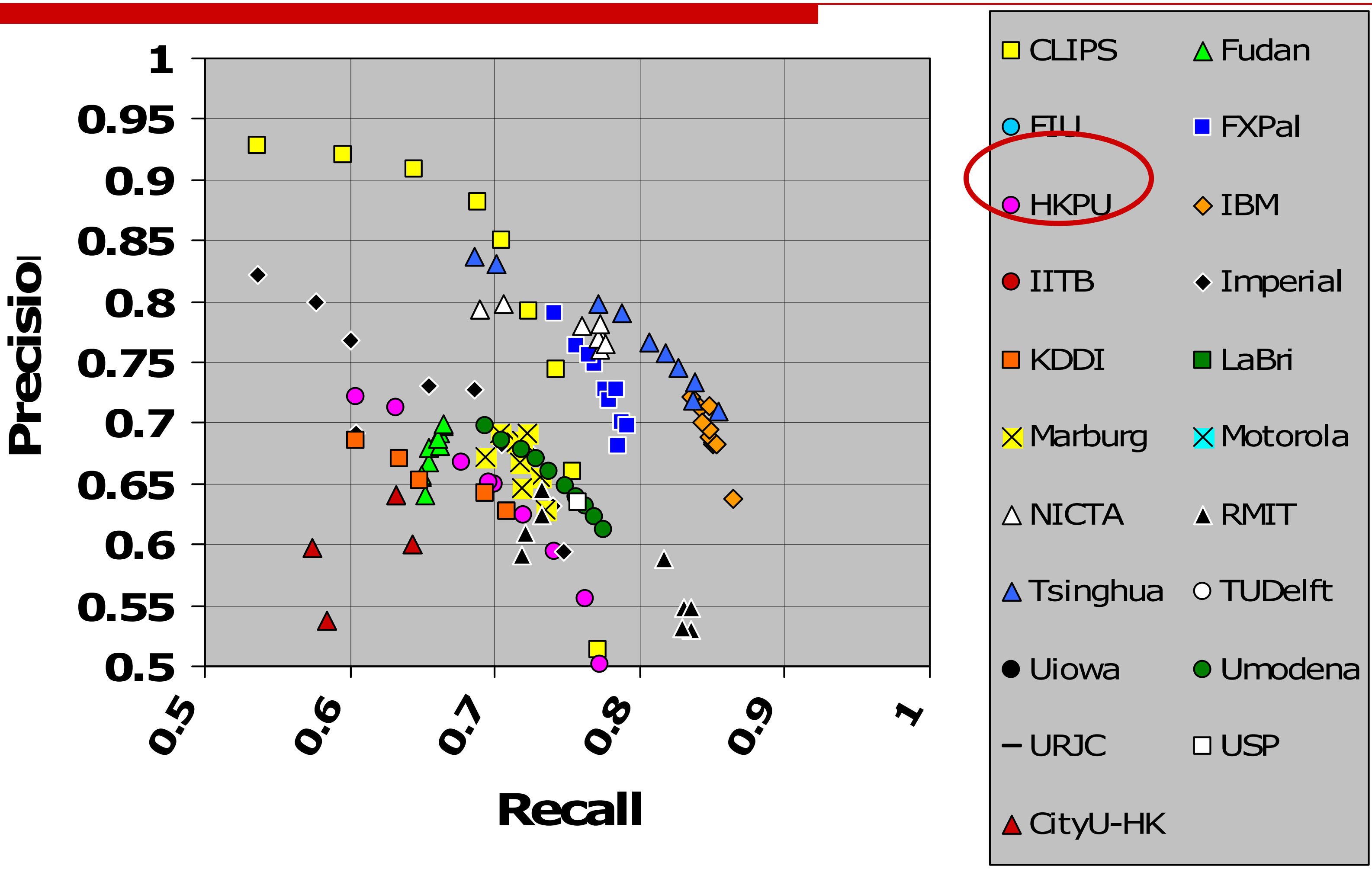
6. Hong Kong Polytechnical University

- Approach
  - Compute frame-frame similarities over different distances and generate distance map;
  - Distance maps have characteristics for cuts, GTs, flashes, etc.
- Performance
  - Computation is about real-time;
- Results

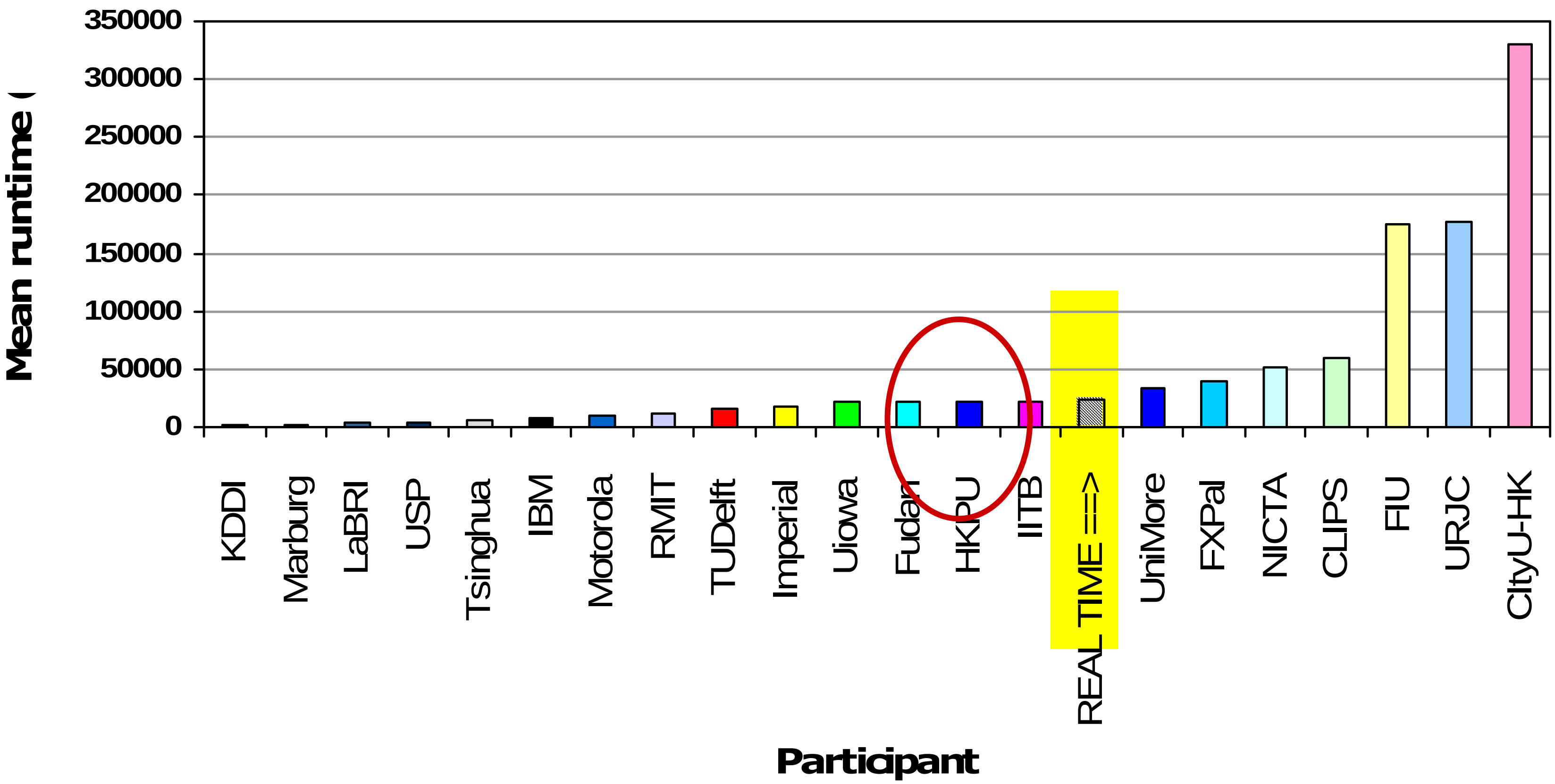
Cuts (zoomed again)



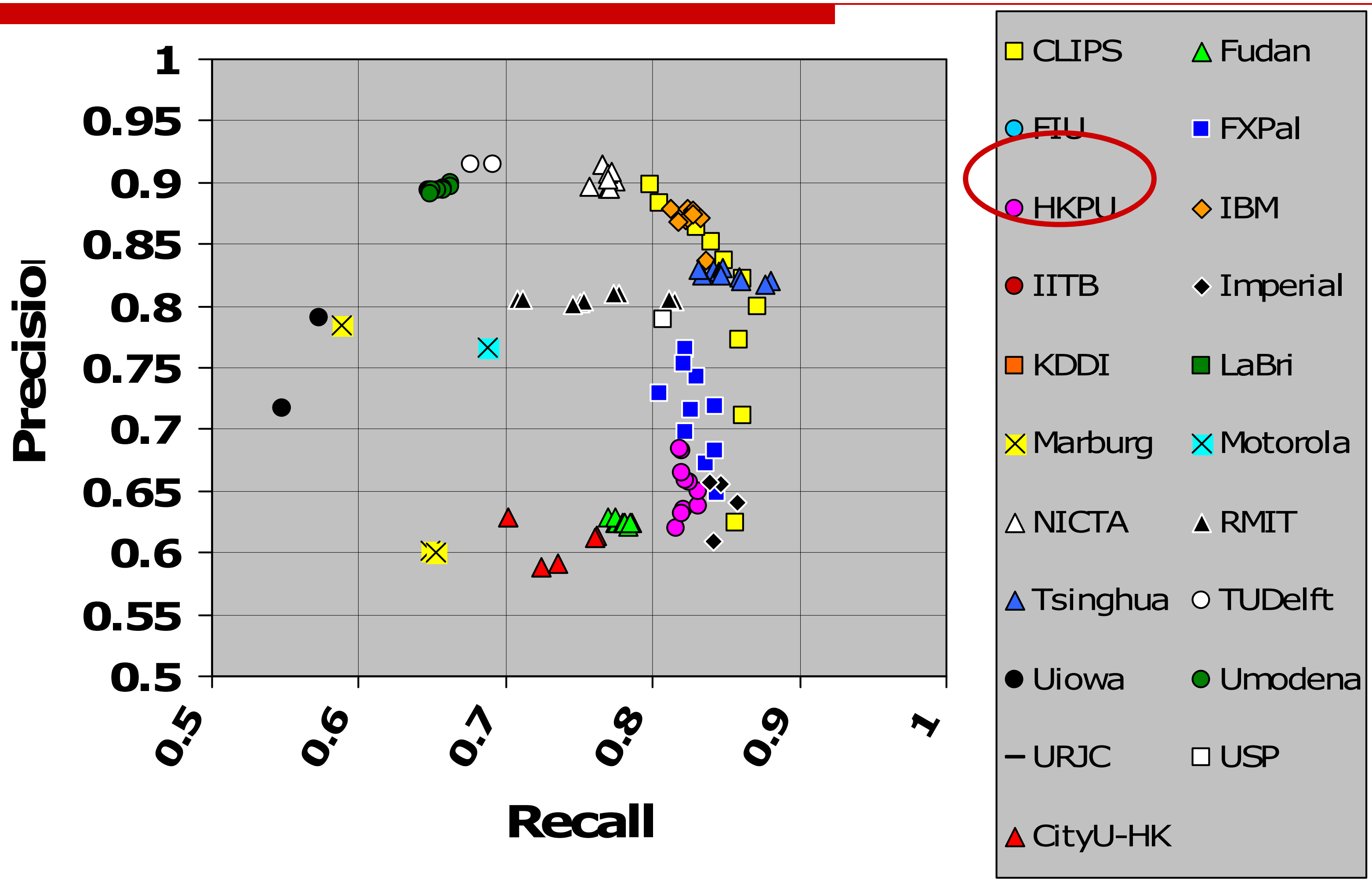
Gradual transitions (zoomed)



Mean runtime in seconds



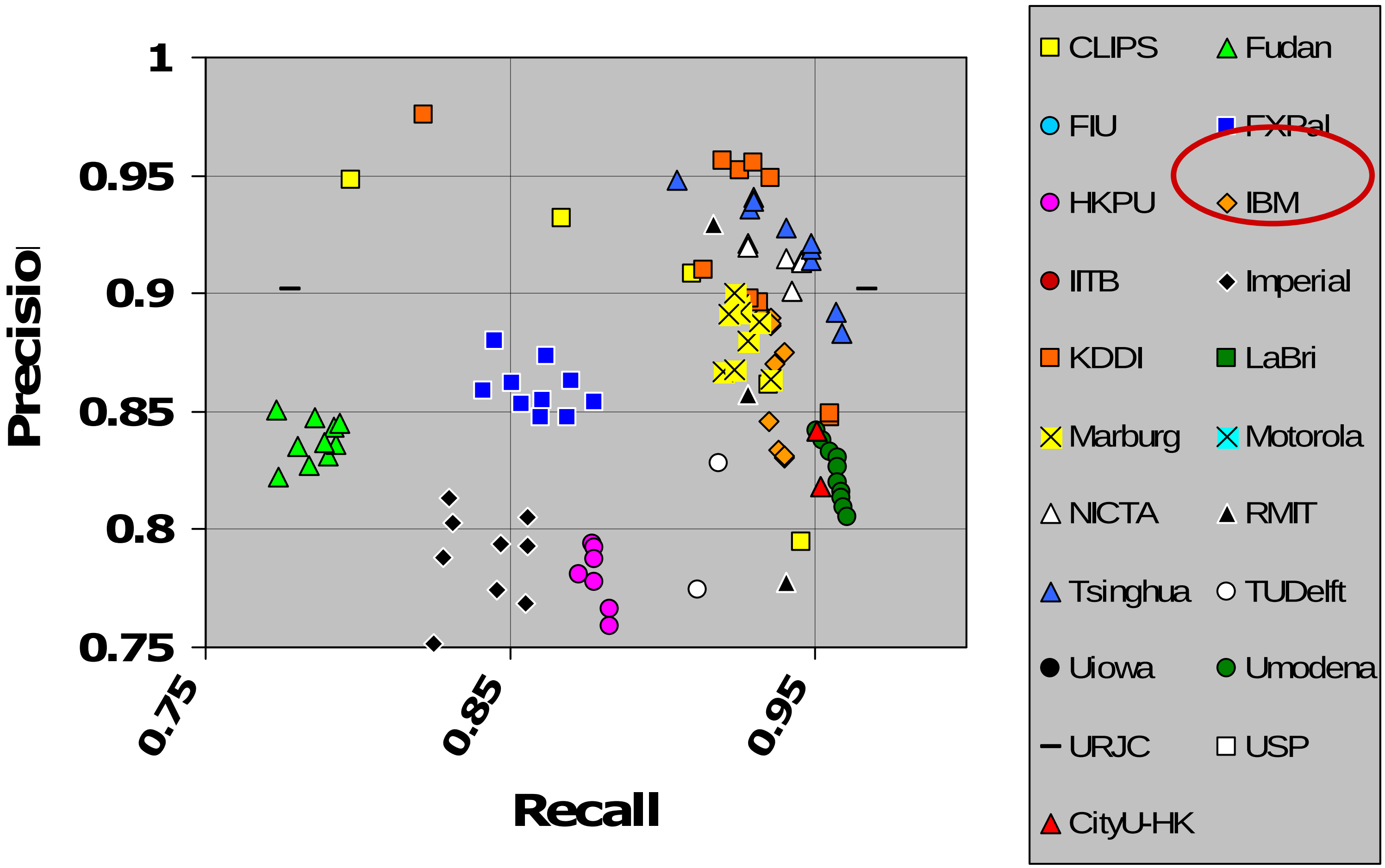
Gradual transitions: Frame-P & R (zoomed)



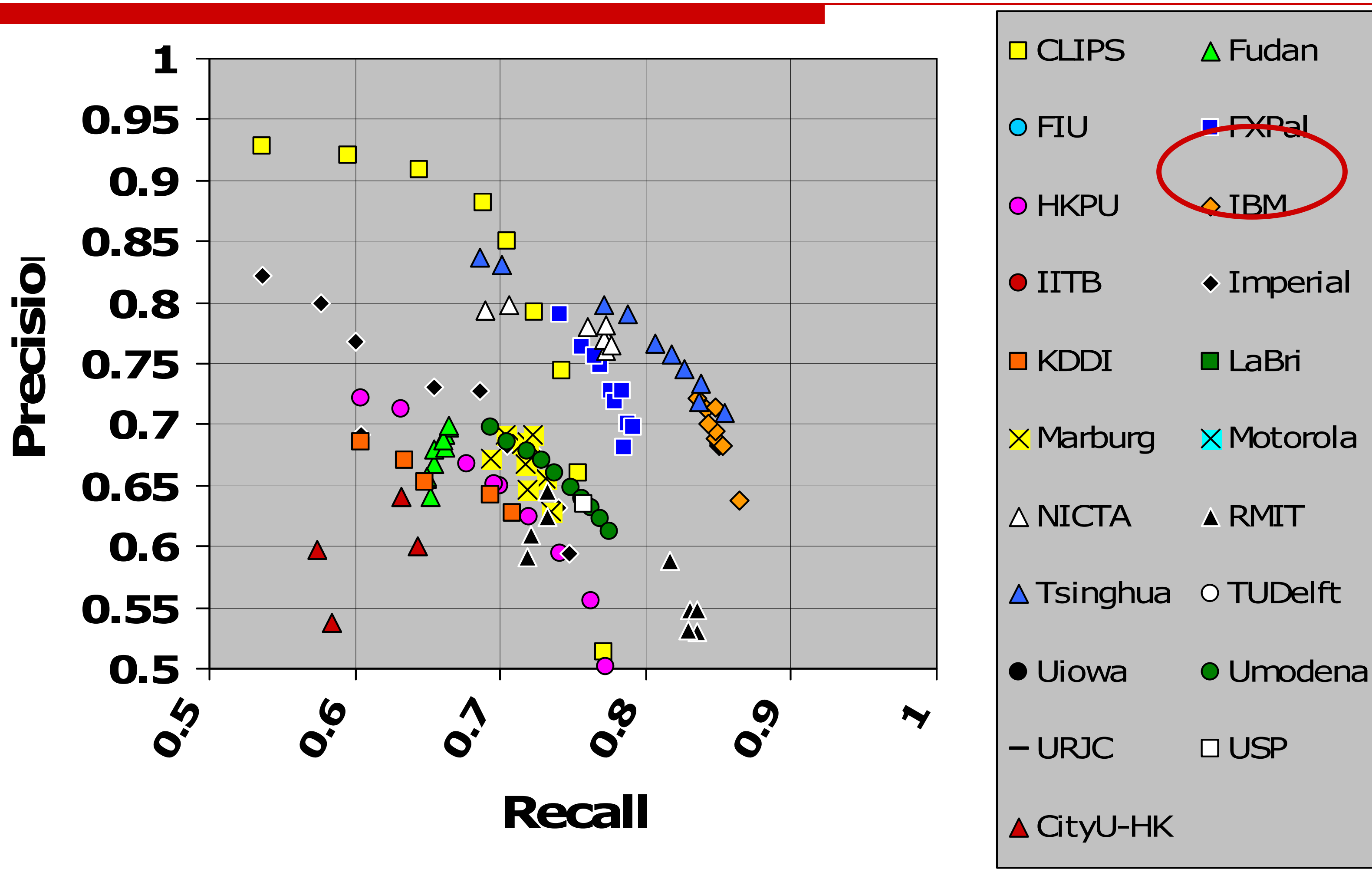
7. IBM Research

- Approach
  - Builds upon previous CueVideo work at TRECVid, system is the same as 2005, except …
  - Noticed that GOP I/P- frame patterns (no B-frames) in TRECVid 2005 video encoding had no B-frames;
  - Used a different video decoder to overcome colour errors;
- Performance
  - Switching the video decoder yielded improved performances;
- Results

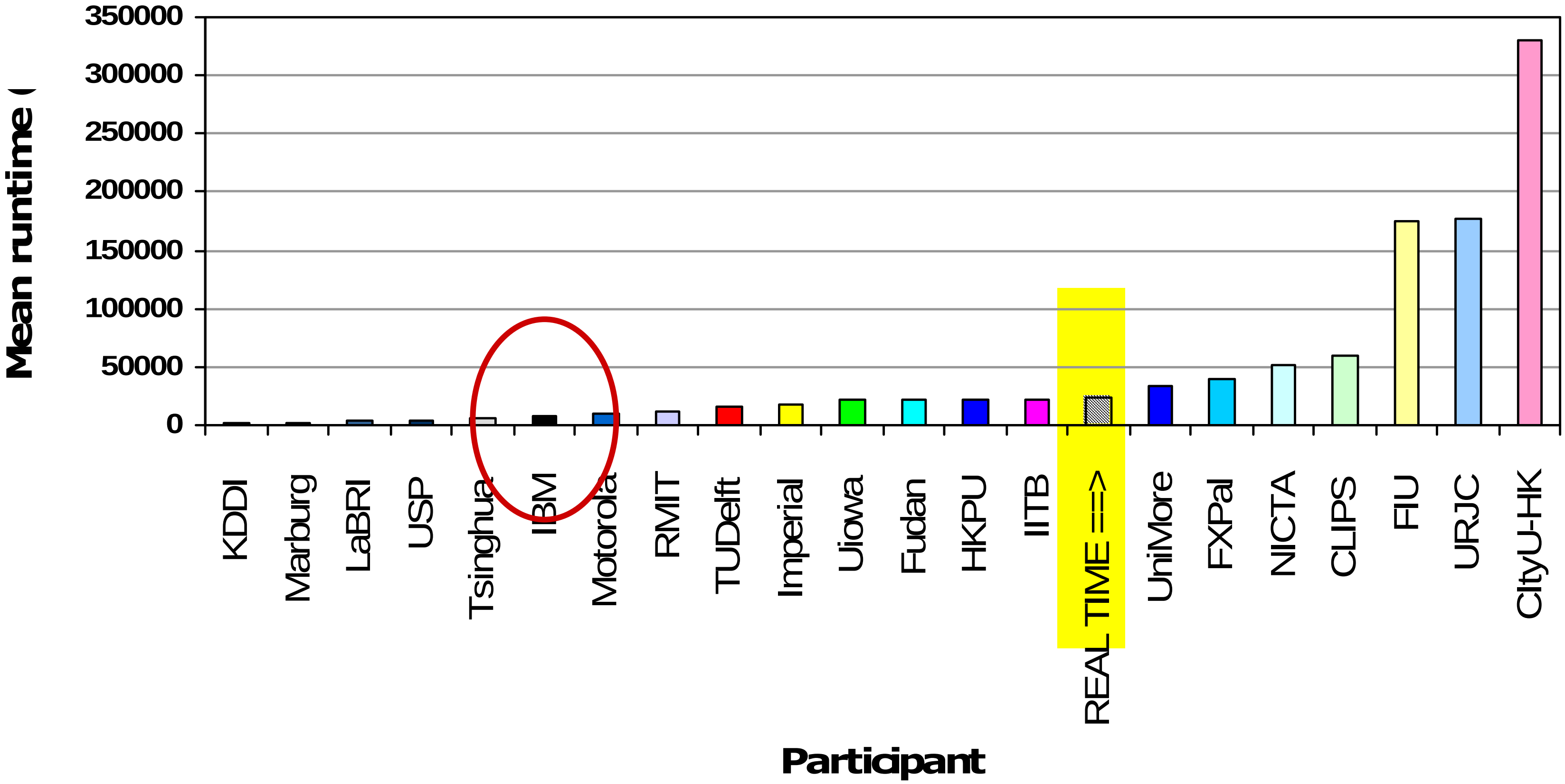
Cuts (zoomed again)



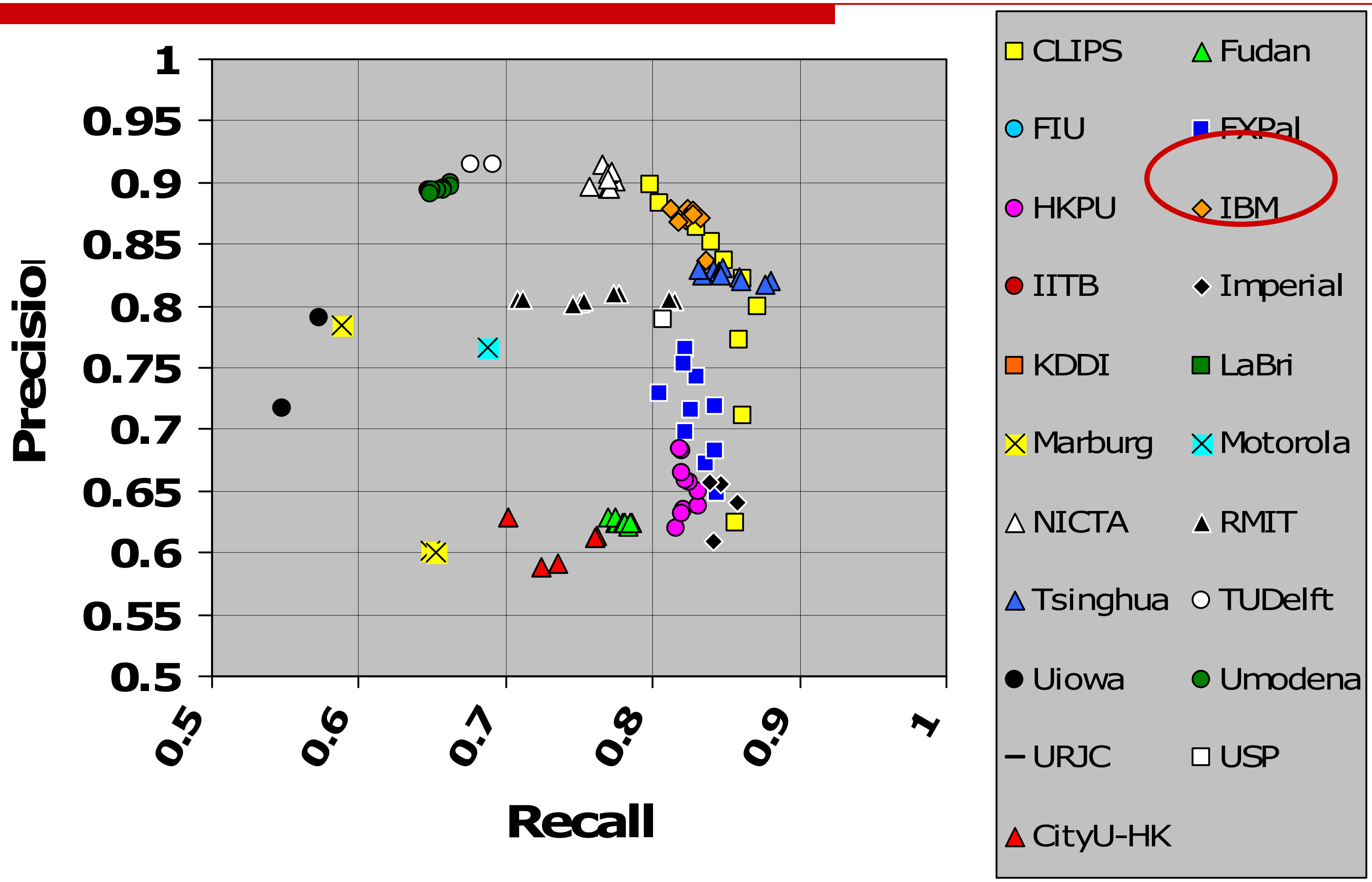
Gradual transitions (zoomed)



Mean runtime in seconds



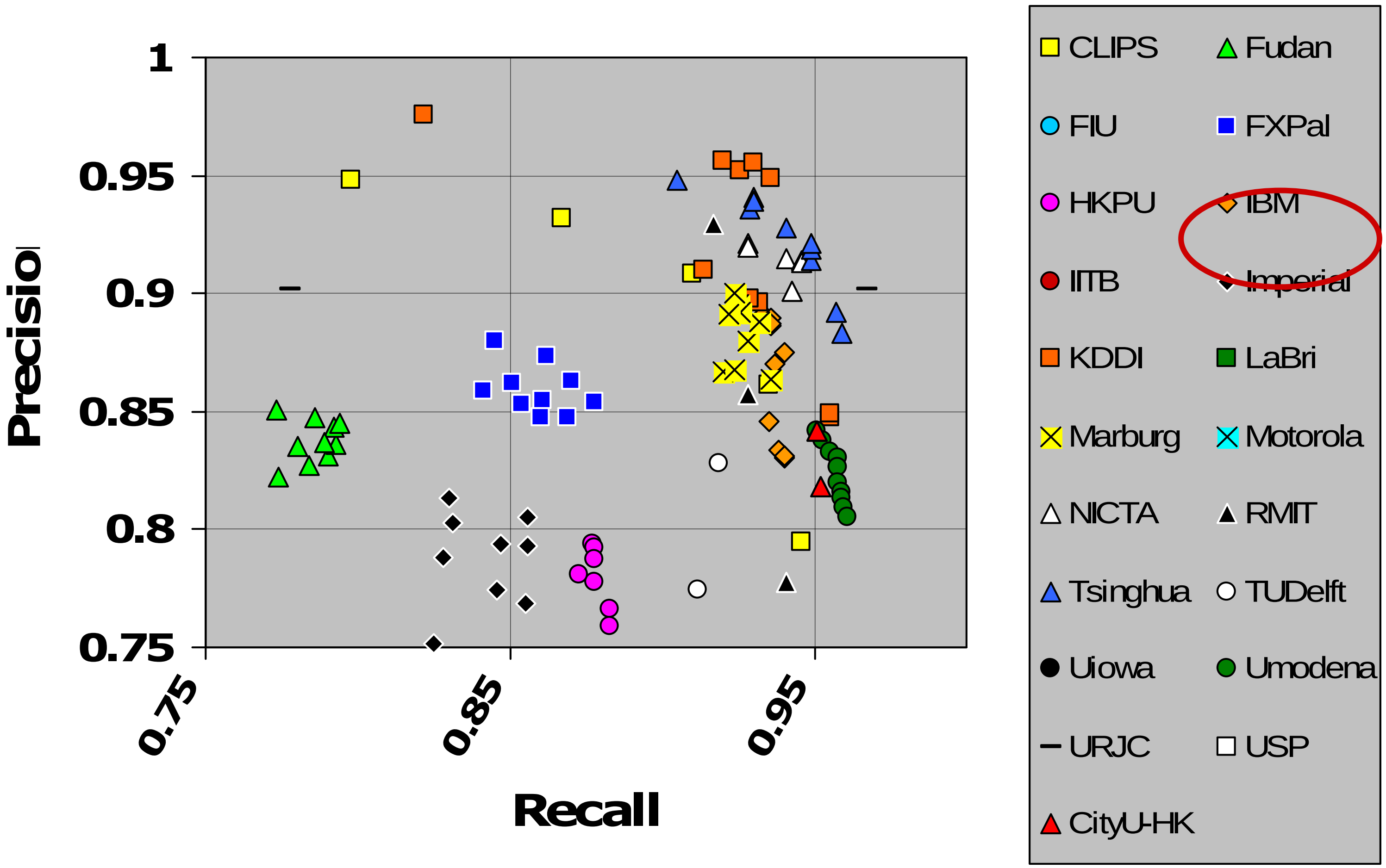
Gradual transitions: Frame-P & R (zoomed)



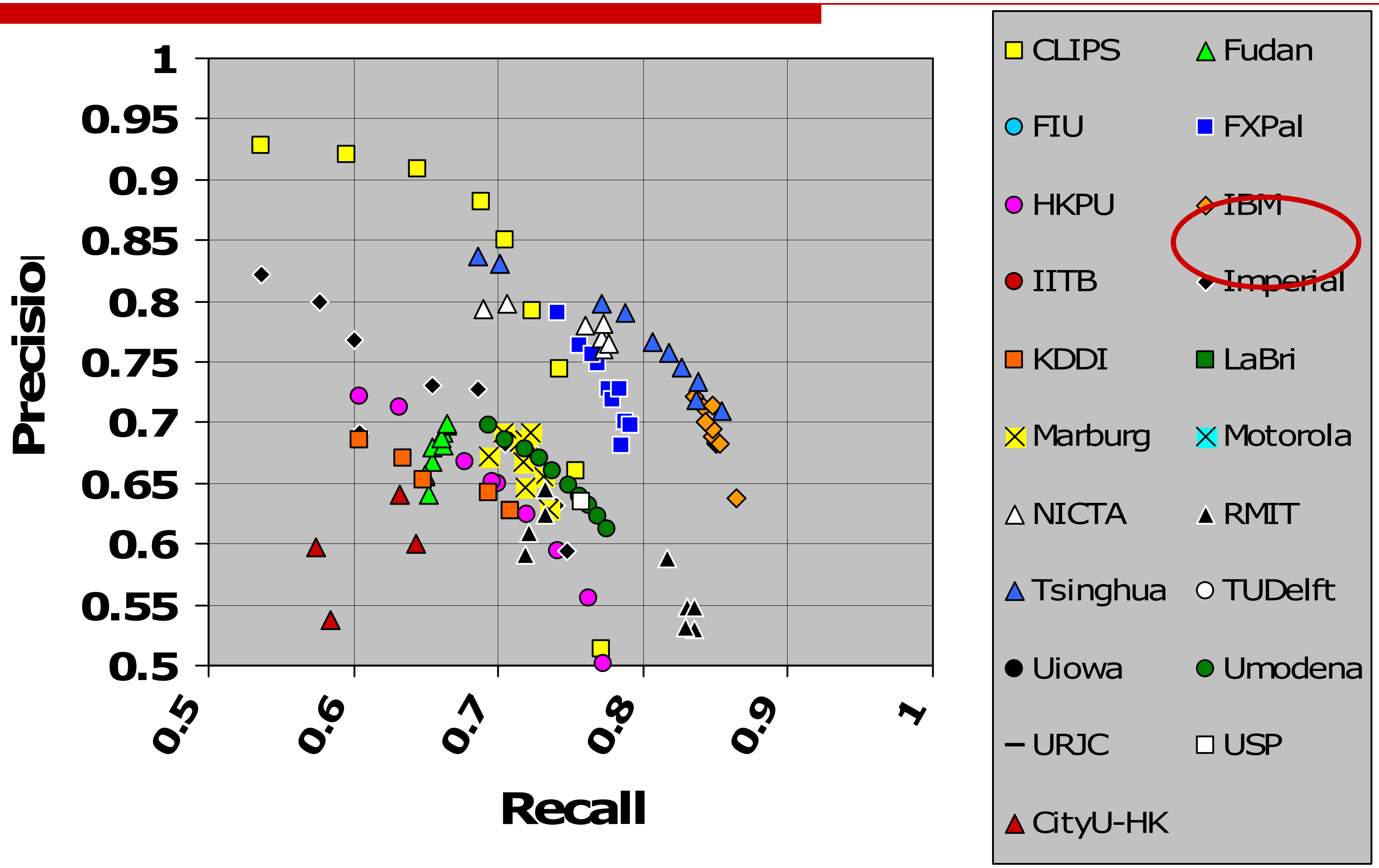
8. Imperial College London

- Approach
  - Same as previous TRECVid submissions;
- Features
  - Exploits frame-frame differences based on colour histogram comparisons
- Results

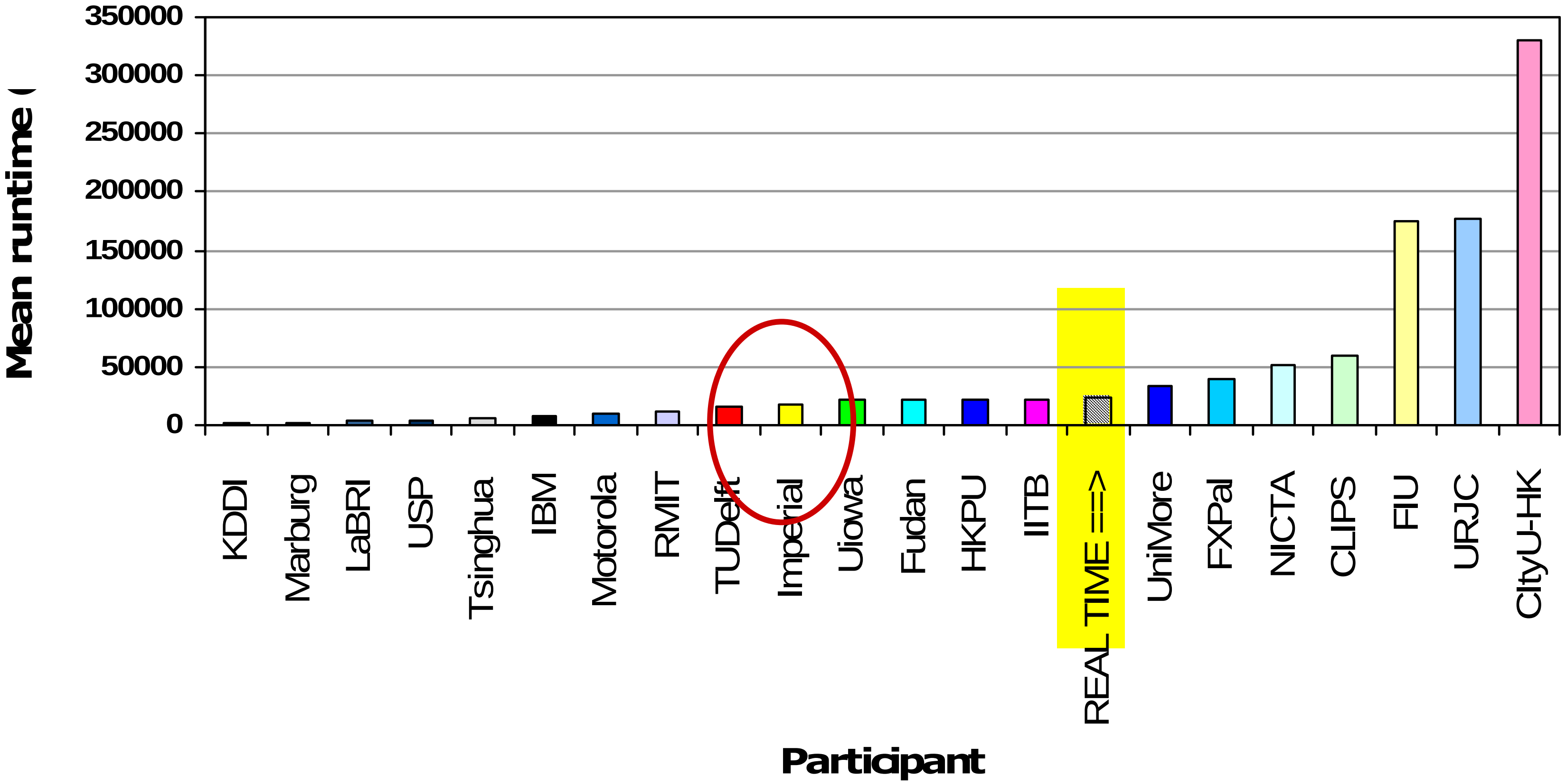
Cuts (zoomed again)



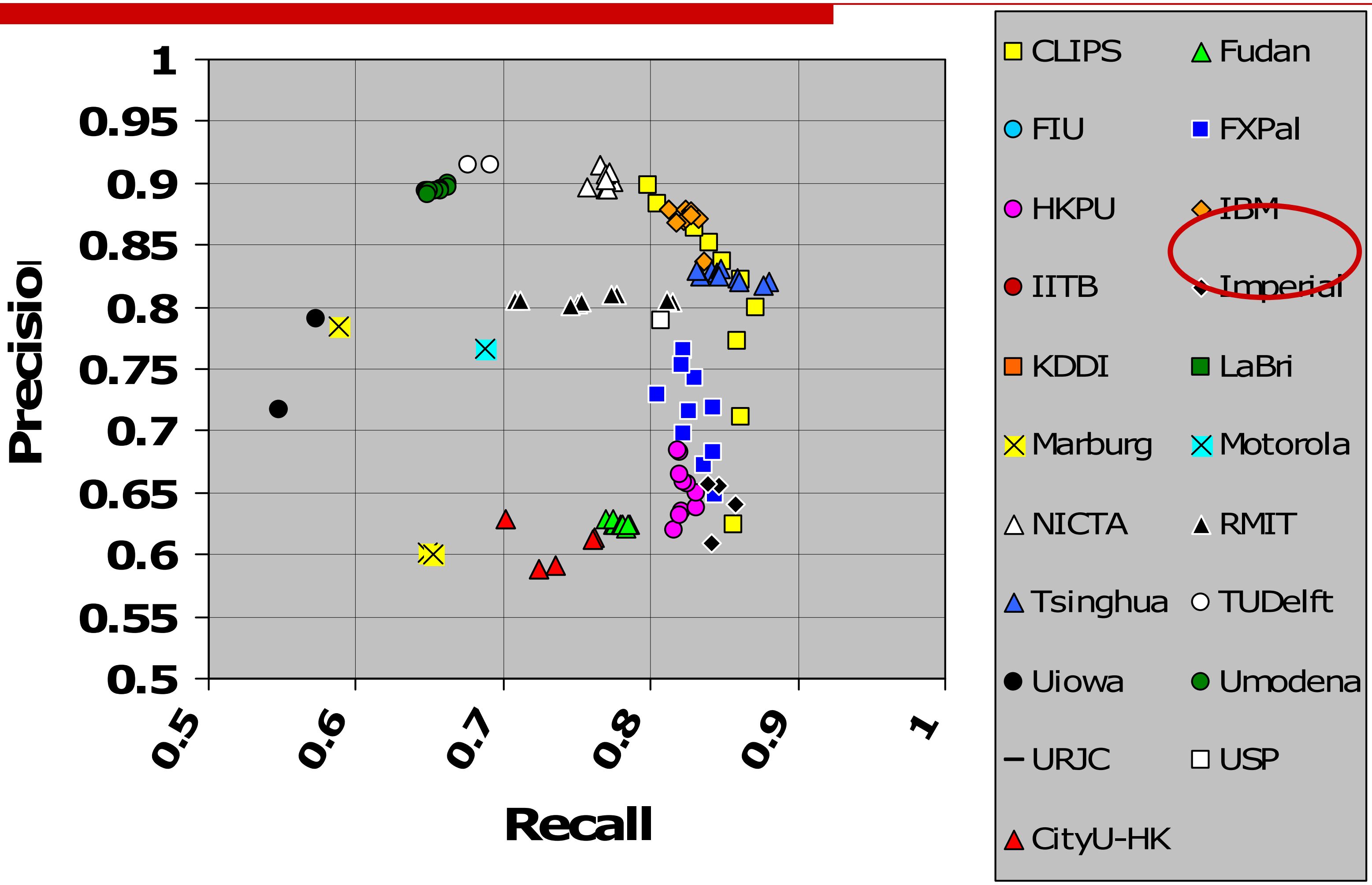
Gradual transitions (zoomed)



Mean runtime in seconds



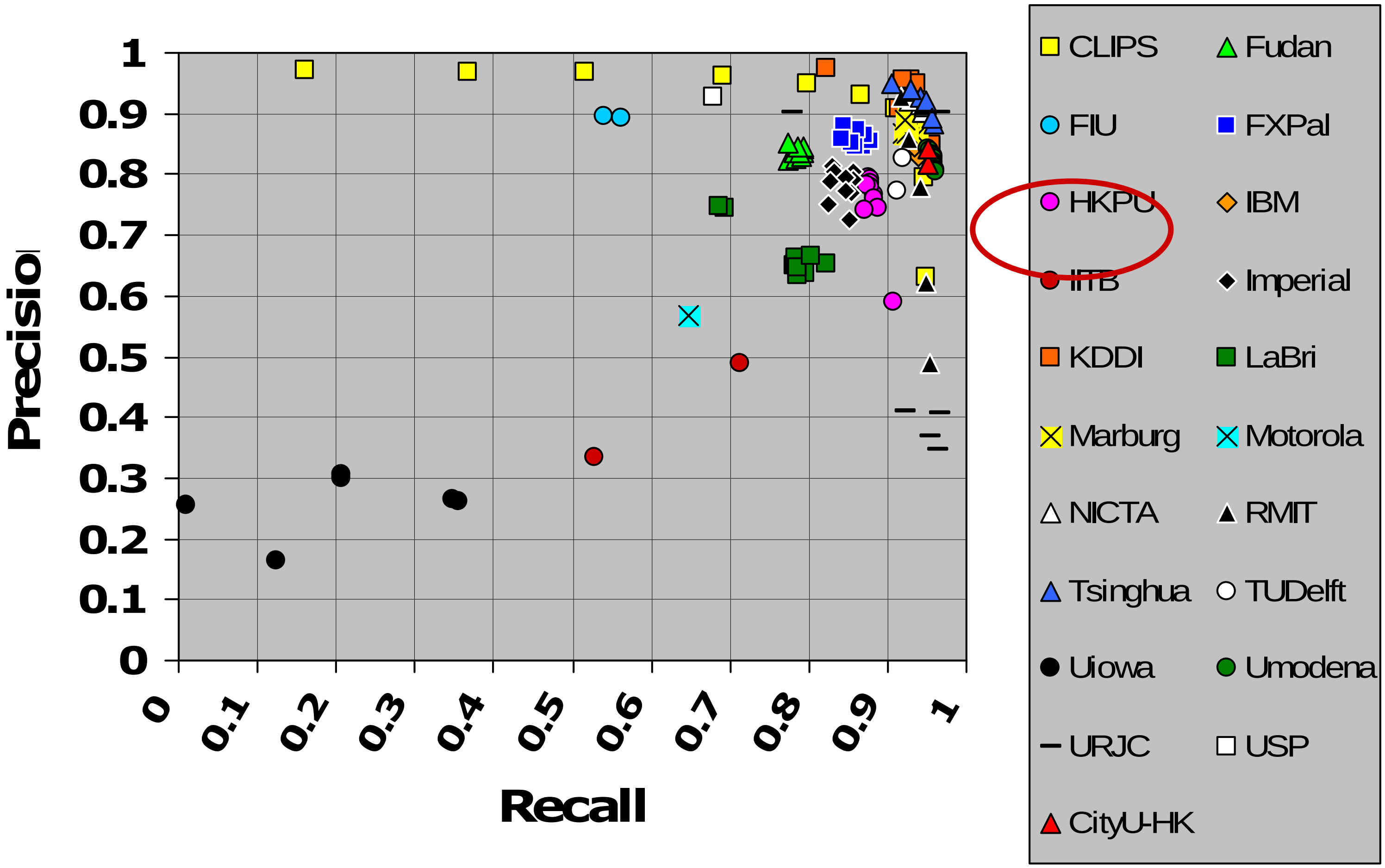
Gradual transitions: Frame-P & R (zoomed)



9. Indian Institute of Technology

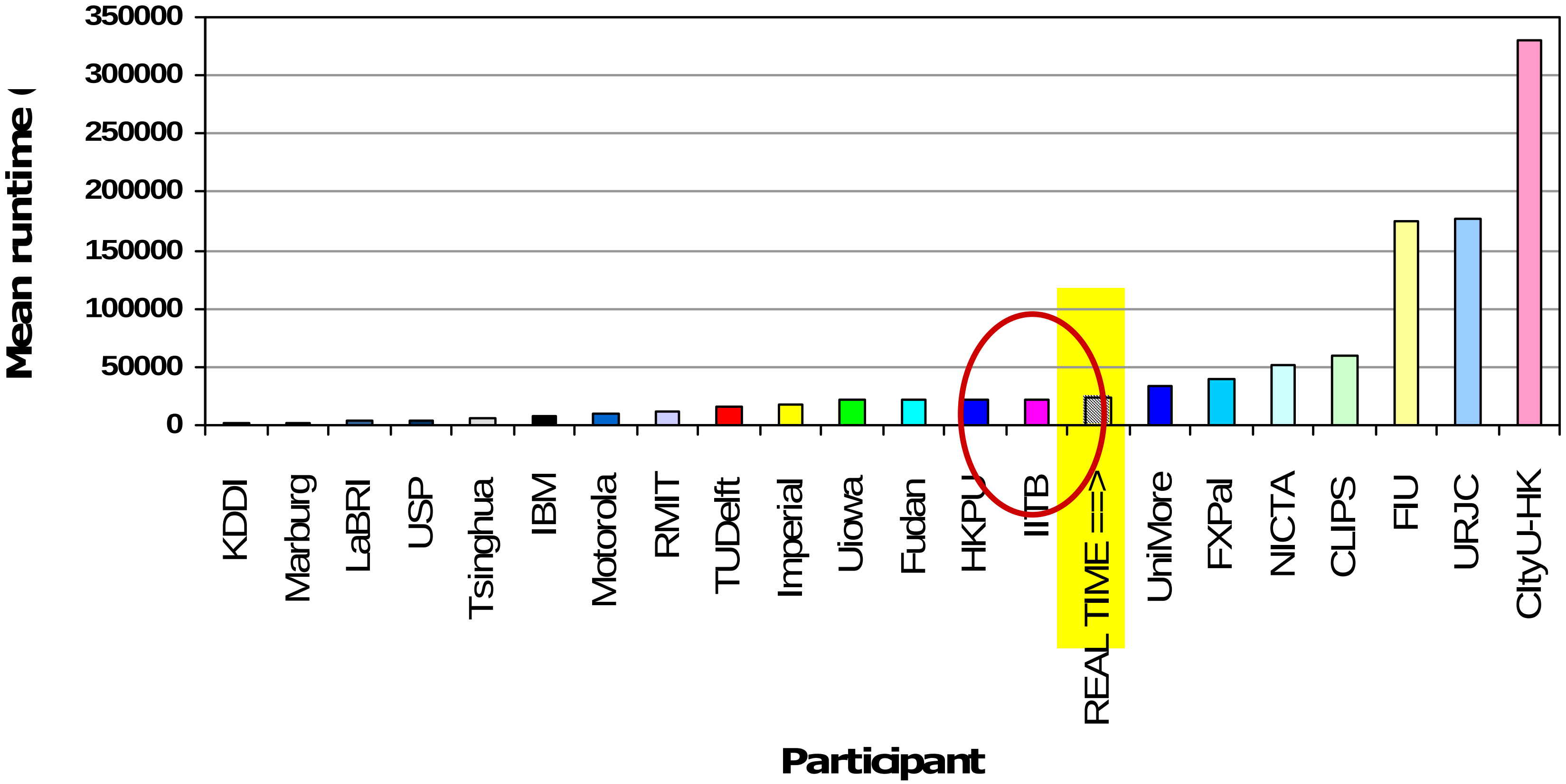
- Approach
  - Addressed false positives caused by abnormal lighting (flashes, reflections, camera movements, explosions, fire, etc.)
- Features
  - 2-pass algorithm - firstly compute similarity between adjacent frames using wavelets, then focus on candidate areas to eliminate false positives;
- Performance
  - Computation about real-time;
- Results
  - Submitted only 1 run, focus on hard cuts only;

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
CLIPS
Fudan
FIU
FXPal
HKPU
IBM
IITB
Imperial
KDDI
LaBri
Marburg
Motorola
NICTA
RMIT
Tsinghua
TUDelft
UIowa
Umodena
URJC
USP
CityU-HK

Mean runtime in seconds

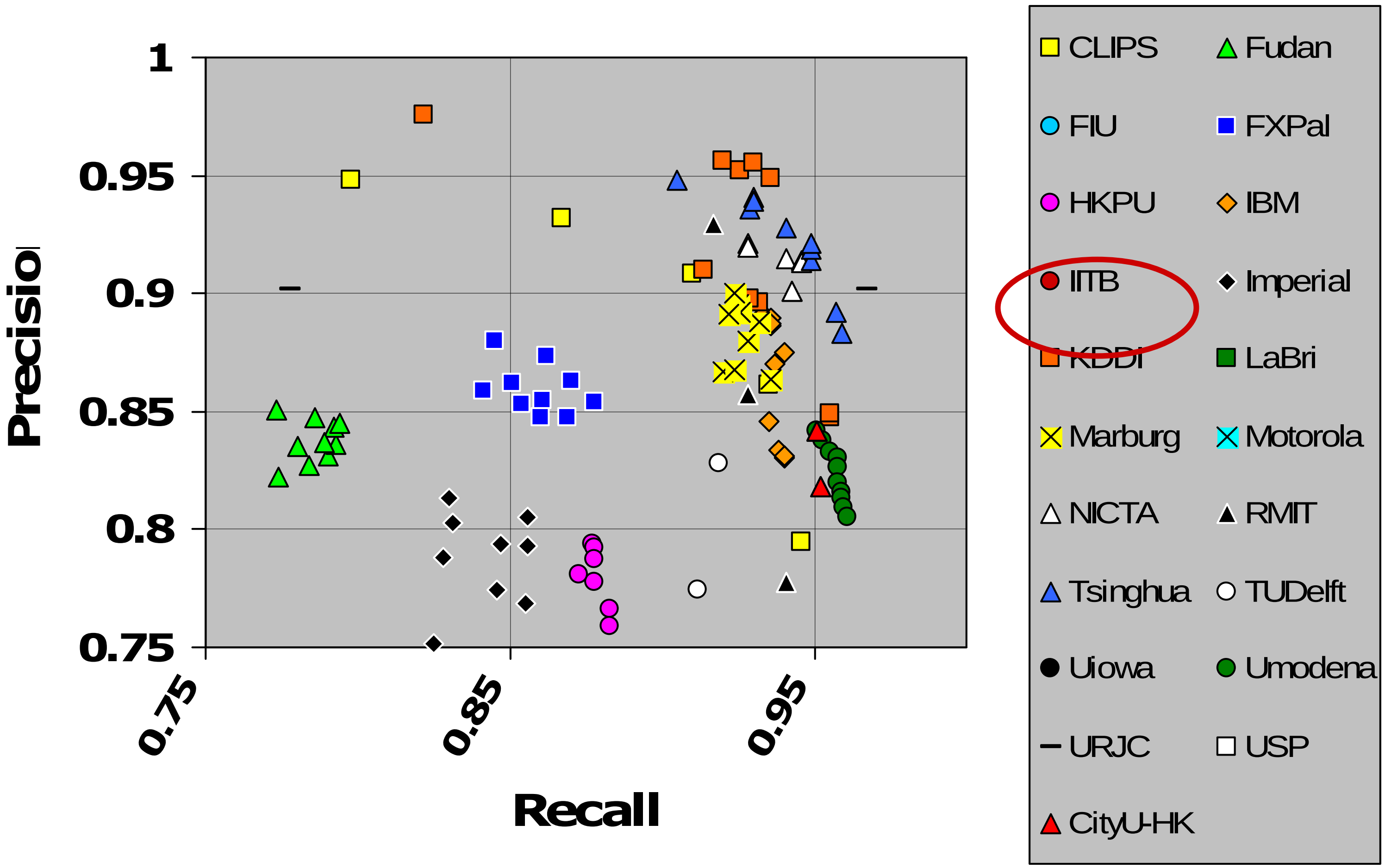


10. KDDI R&D Laboratories, Inc.

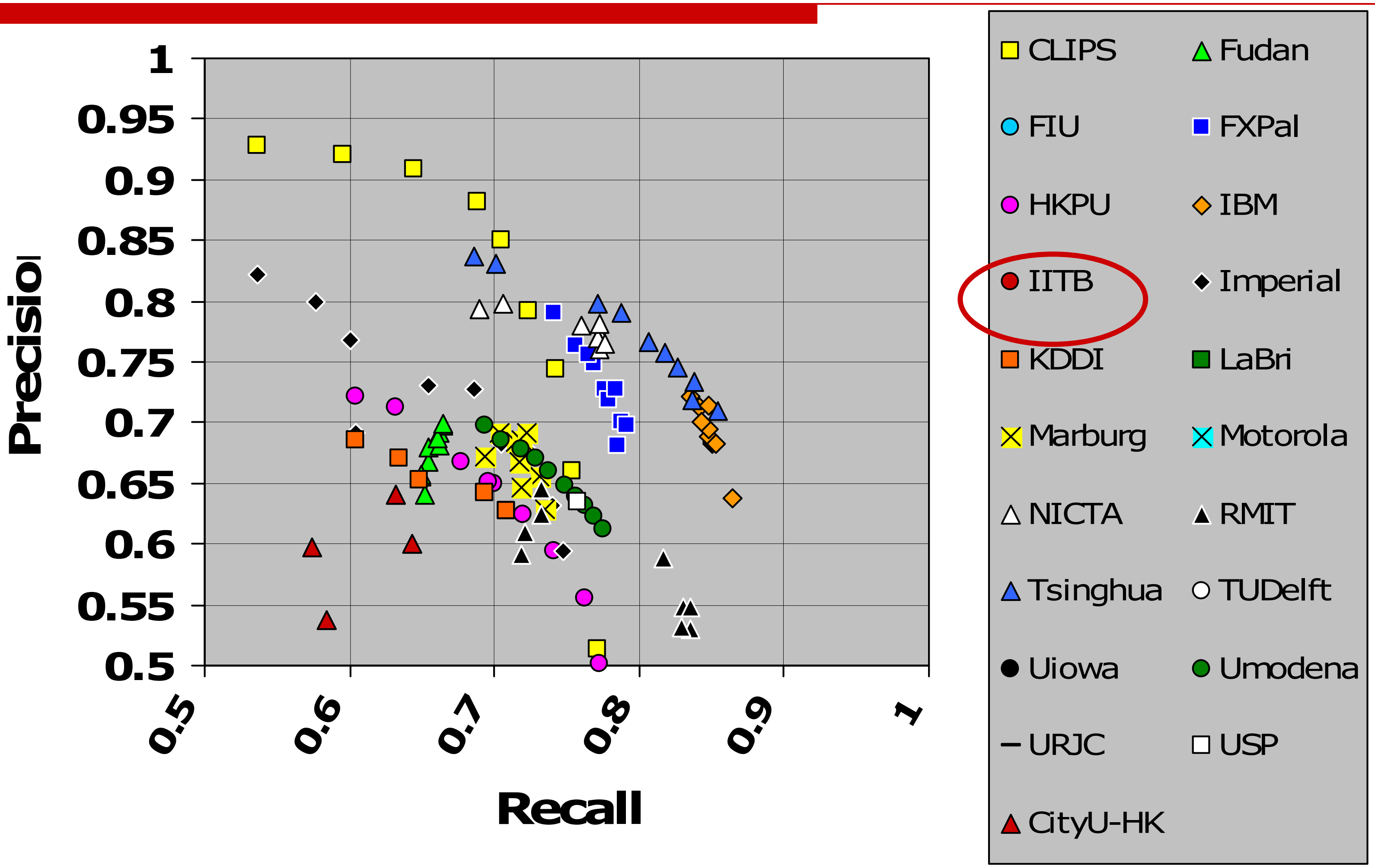
- Approach
  - Late arrival of paper - available at registration desk;
  - Compressed domain - hence fast;
  - Luminance adaptive threshold and image cropping equals goo results;

  - Last year worked in the compressed domain, extending an approach by adding edge features from DC image, colour layout, and SVM learning;
- Results
  - Worth looking at …

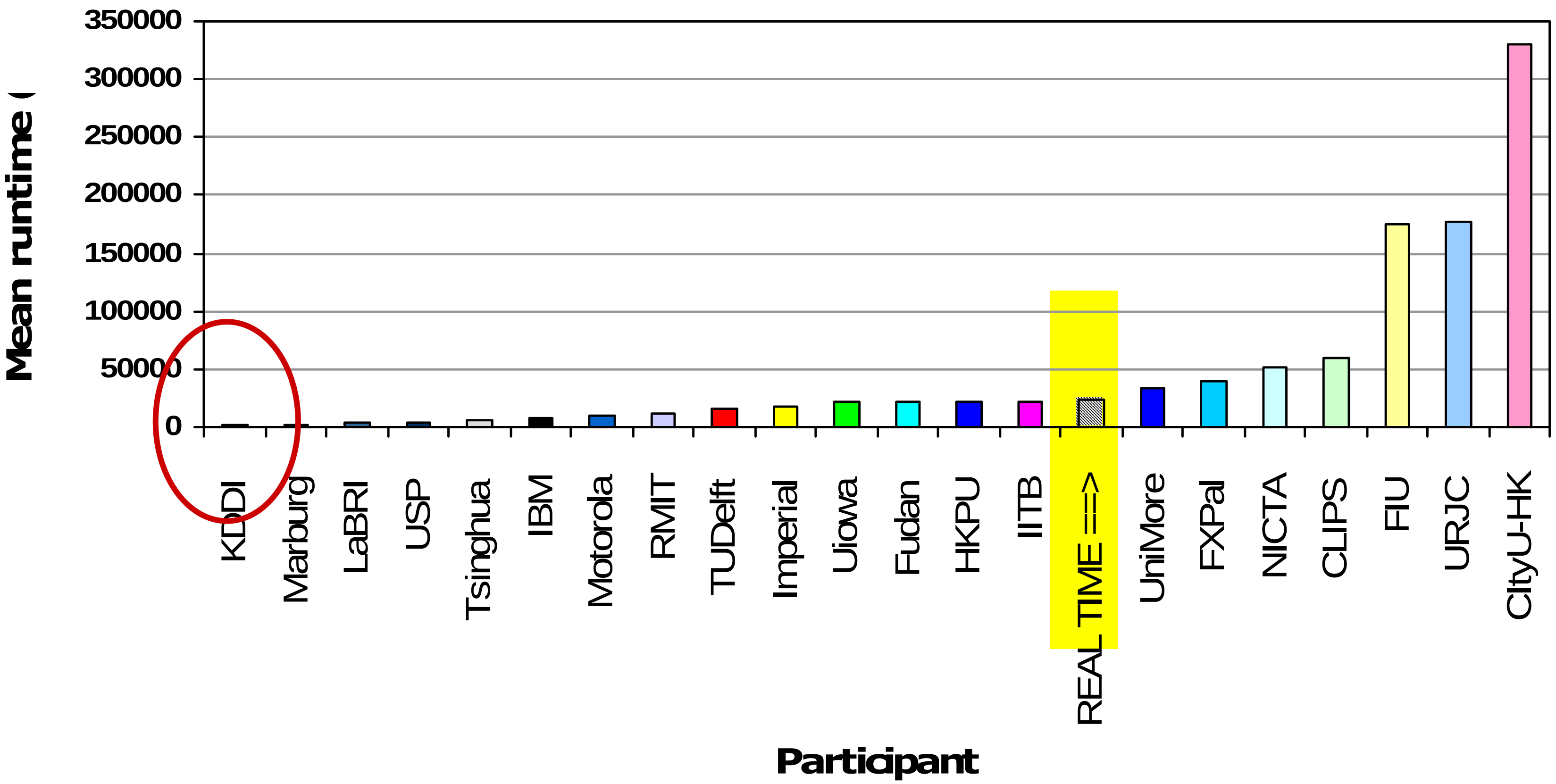
Cuts (zoomed again)



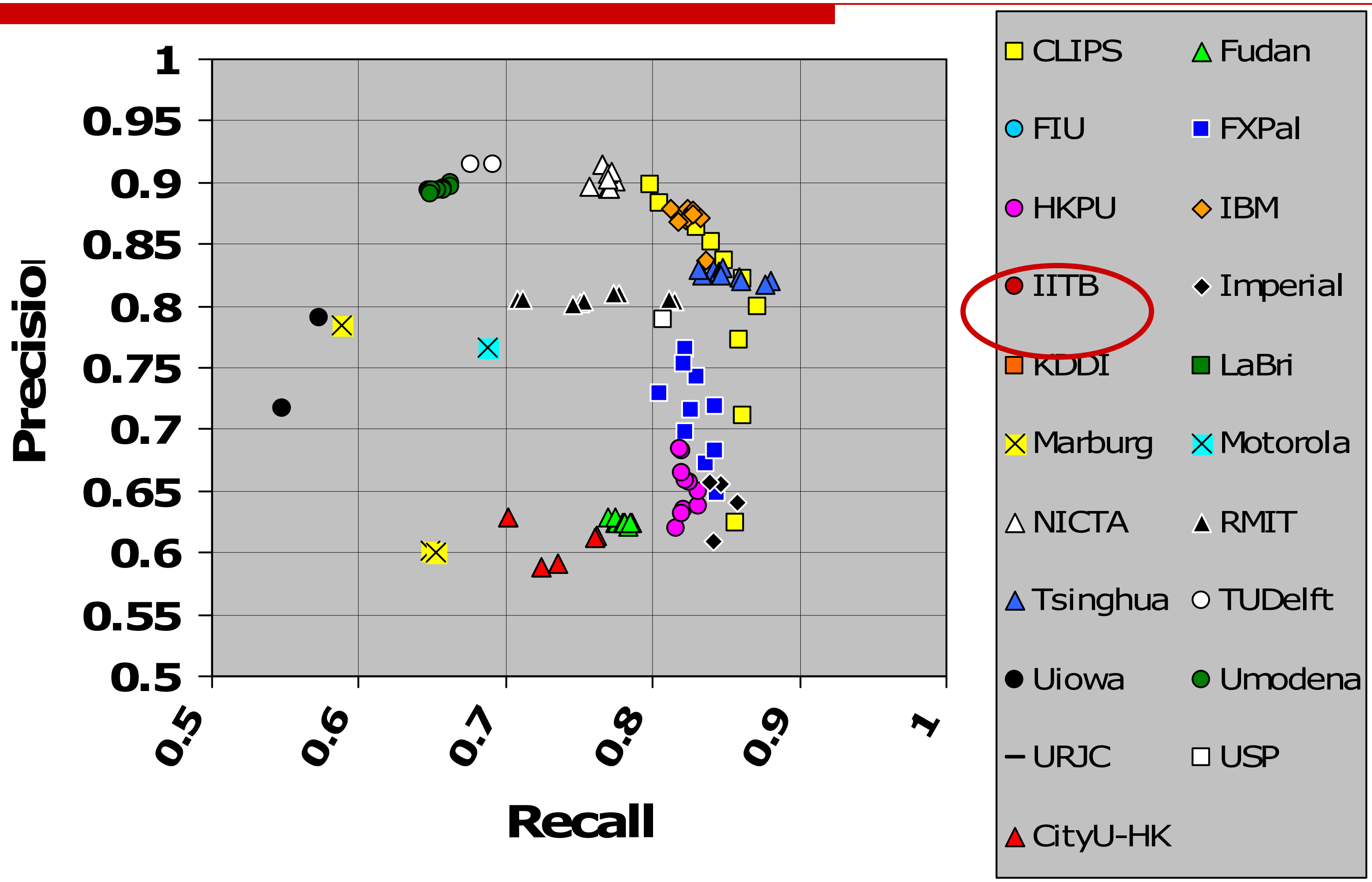
Gradual transitions (zoomed)



Mean runtime in seconds



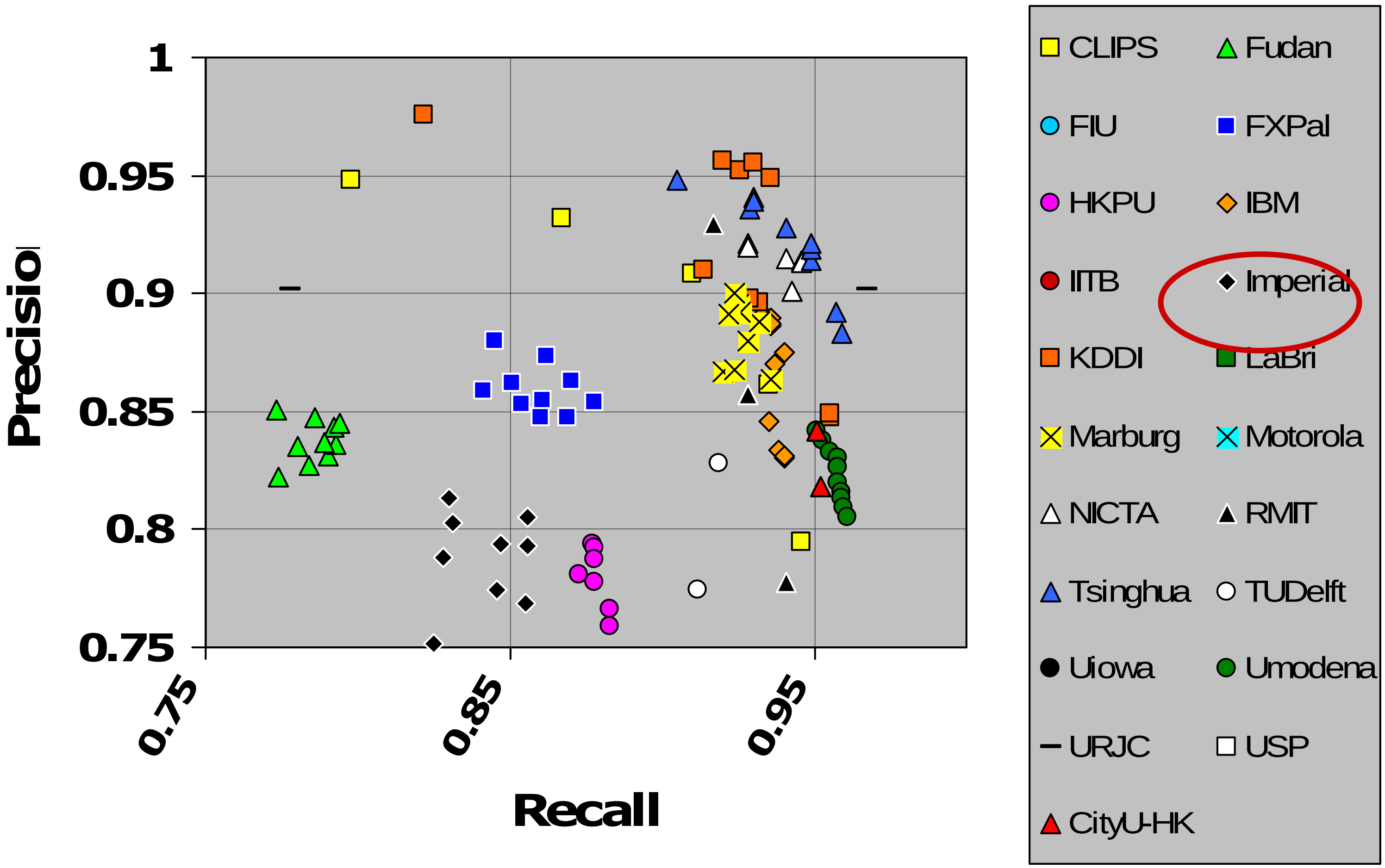
Gradual transitions: Frame-P & R (zoomed)



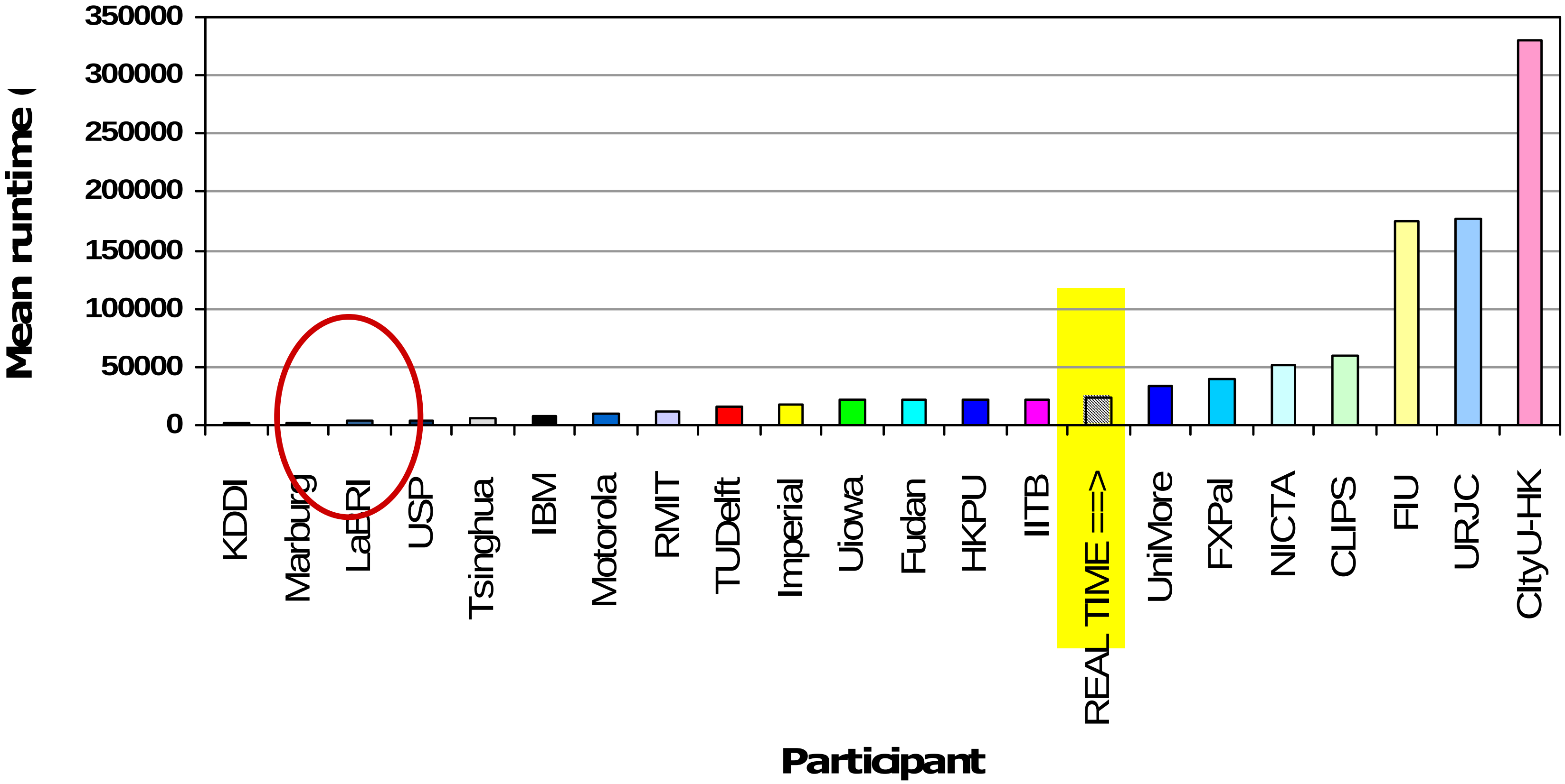
11. LaBRI

- Approach
  - Last year worked in compressed domain, computing motion and frame statistics, then measure similarity between compensated adjacent I-frames;
  - This year most effort in camera motion task but submitted SBD runs based on this
- Performance
  - Good on hard cuts, and fast, not good on GTs
- Results

Cuts (zoomed again)



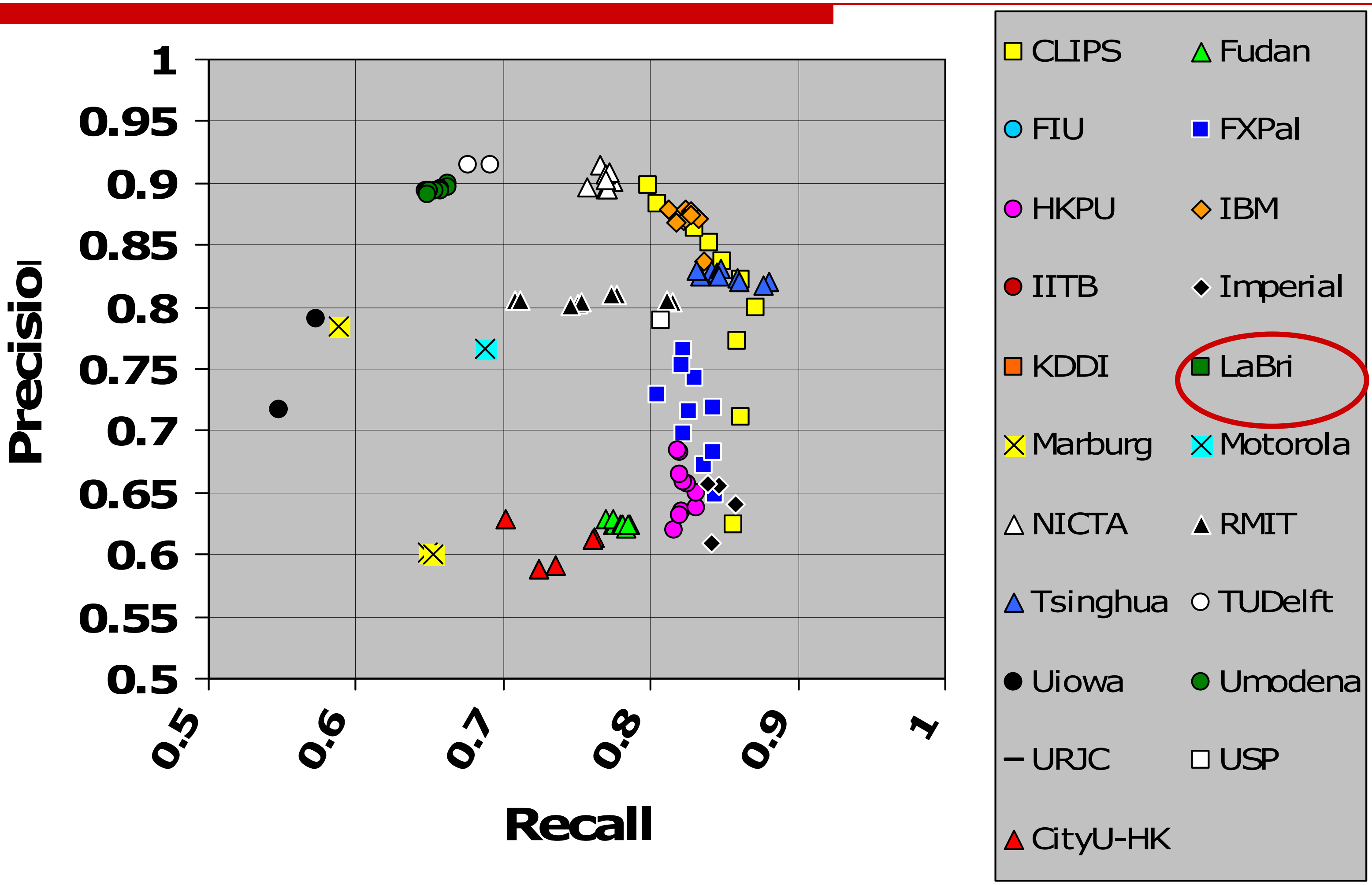
Mean runtime in seconds



12. Motorola Multimedia Research Laboratory

- Approach
  - Didn’t submit a paper so we don’t know !
- Results
  - Fast execution but don’t appear in the zoomed areas of graphs except for …

Gradual transitions: Frame-P & R (zoomed)

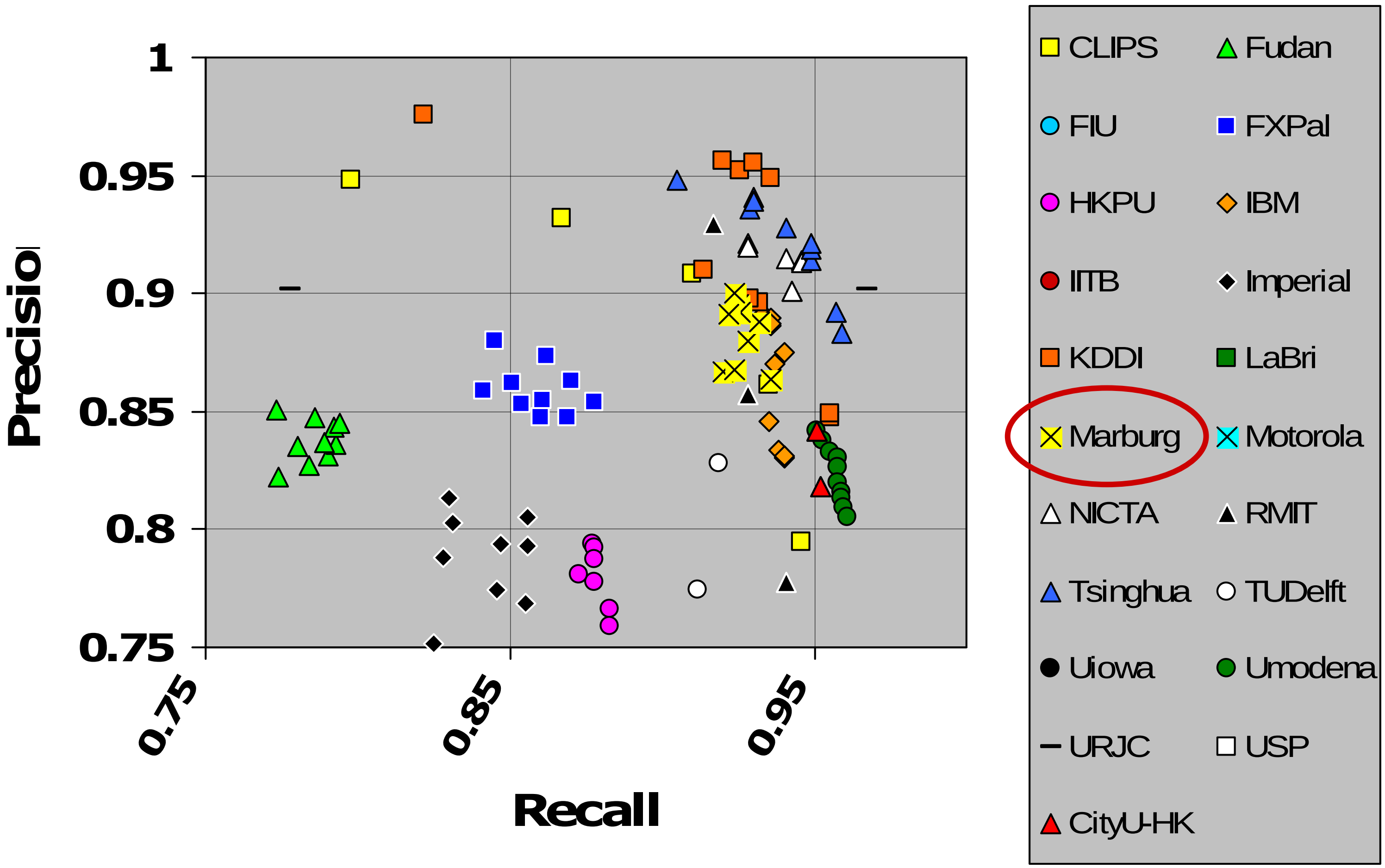


13. National ICT Australia

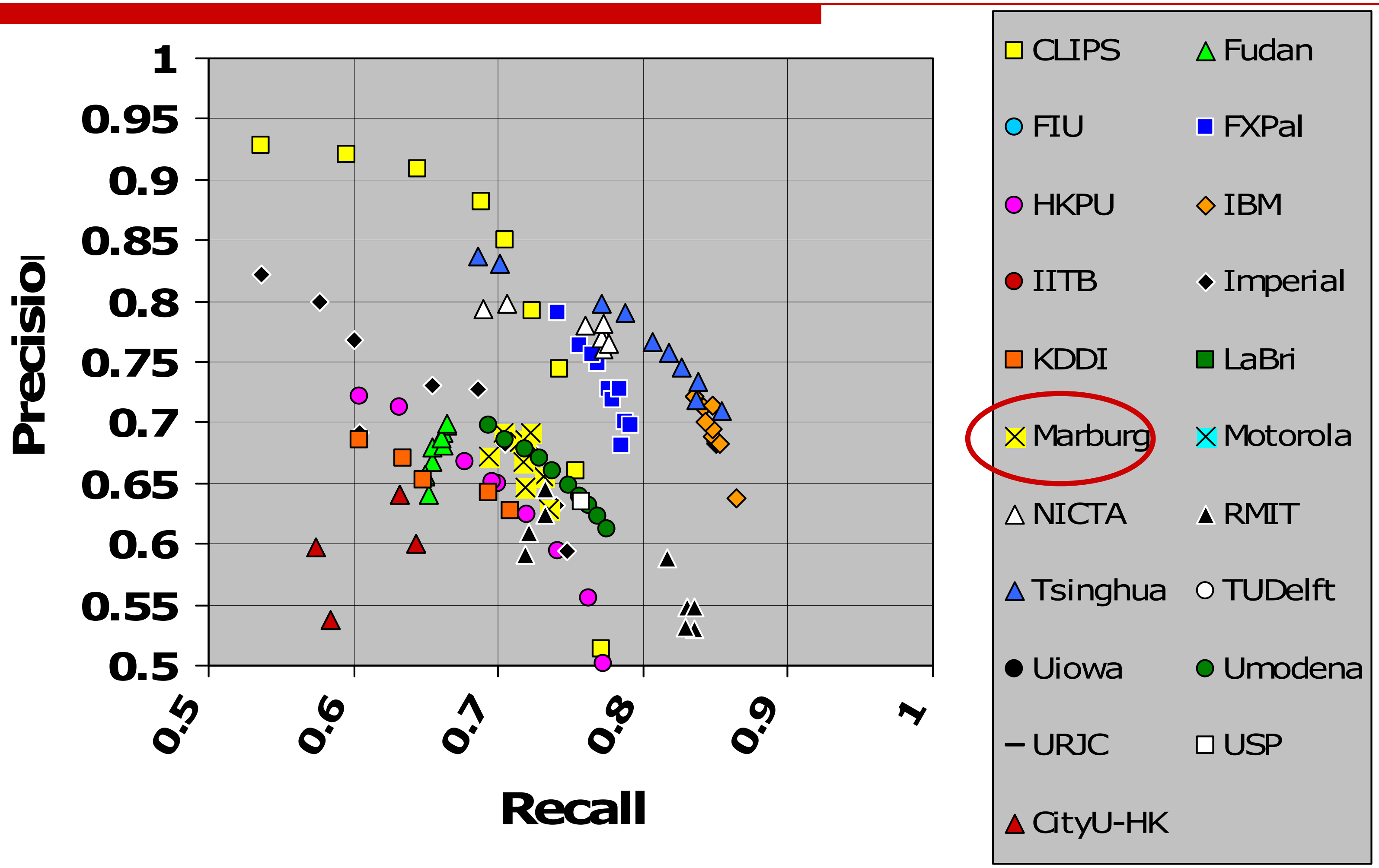
- Approach
  - Late paper submitted and it doesn't reveal much … "Video analysis + machine learning: - New to TRECVID - Developers-- Drs Zhenghua (Jack) Yu, SVN Vishwanathan and Alex Smola"

- Results
  - Expensive computation but worth a peek at …

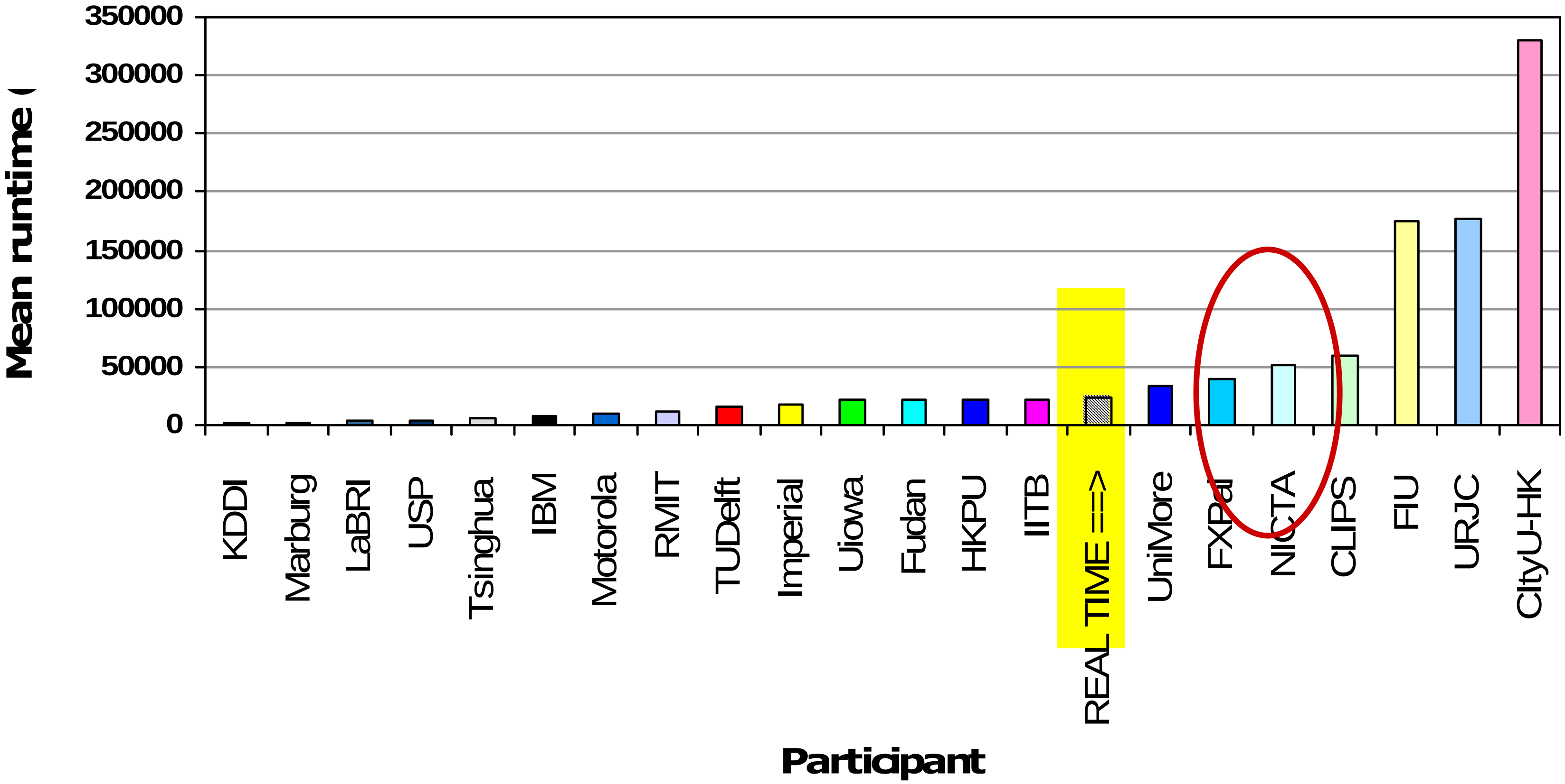
Cuts (zoomed again)



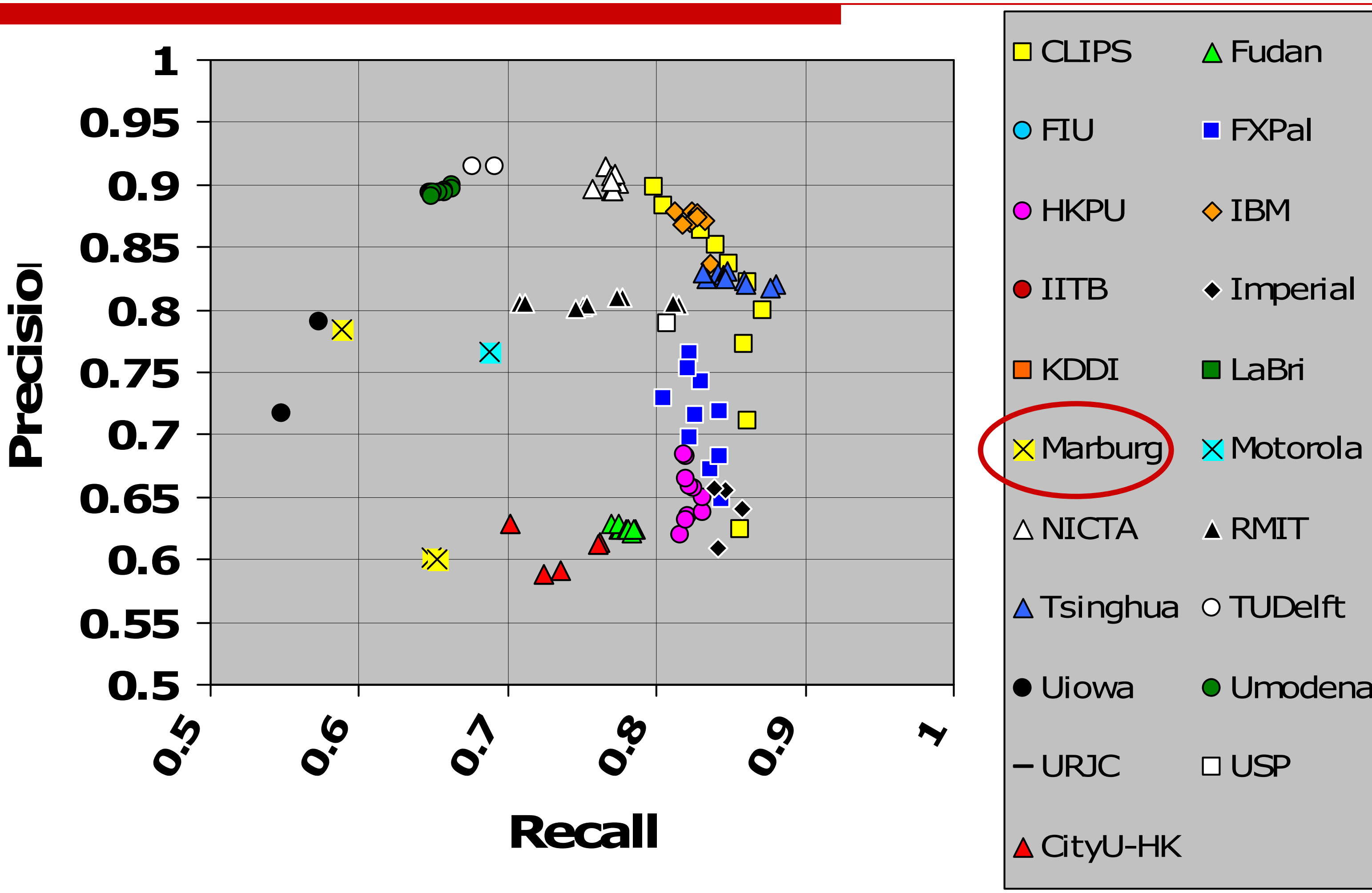
Gradual transitions (zoomed)



Mean runtime in seconds



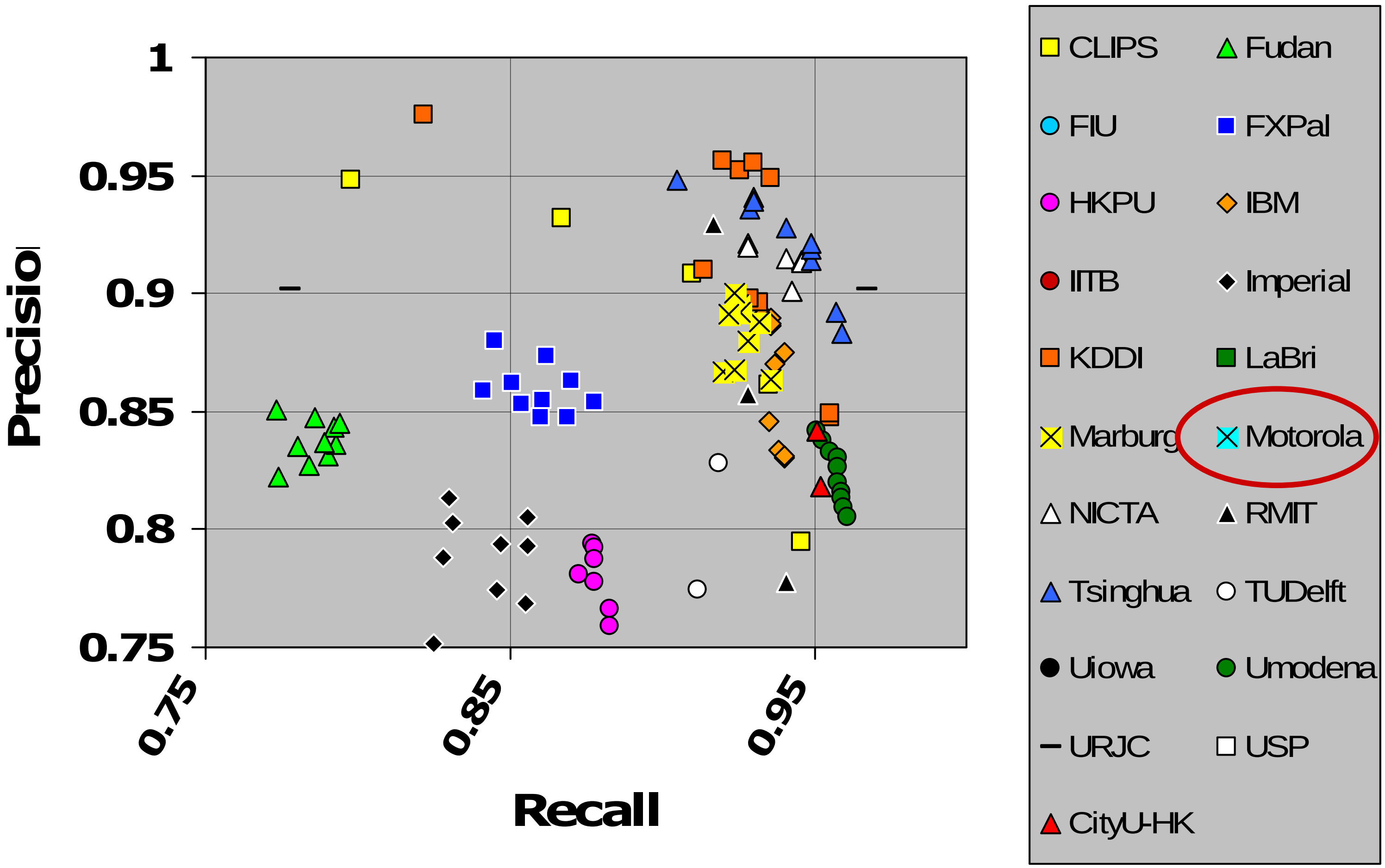
Gradual transitions: Frame-P & R (zoomed)



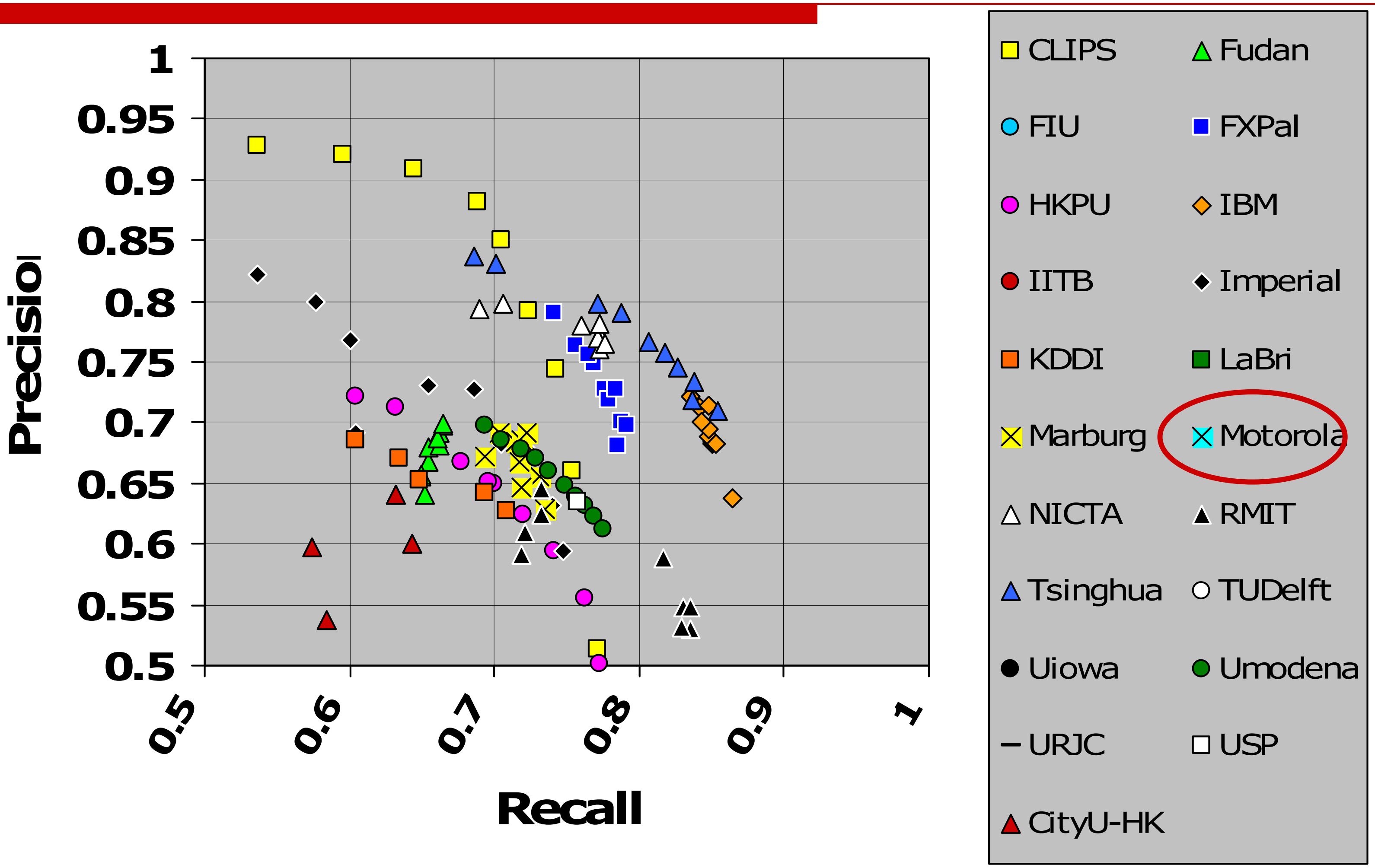
14. RMIT University

- Approach
  - New implementation of their sliding query window approach, compute frame similarities among X frames before/after;
  - Frame similarities based on colour histograms;
  - Experimented with different (HSV) colour histogram representations;
- Features
  - Feature selection/reduction yielded improved performances;
- Performance
  - Not as good as expected because sensitive to training data;
- Results

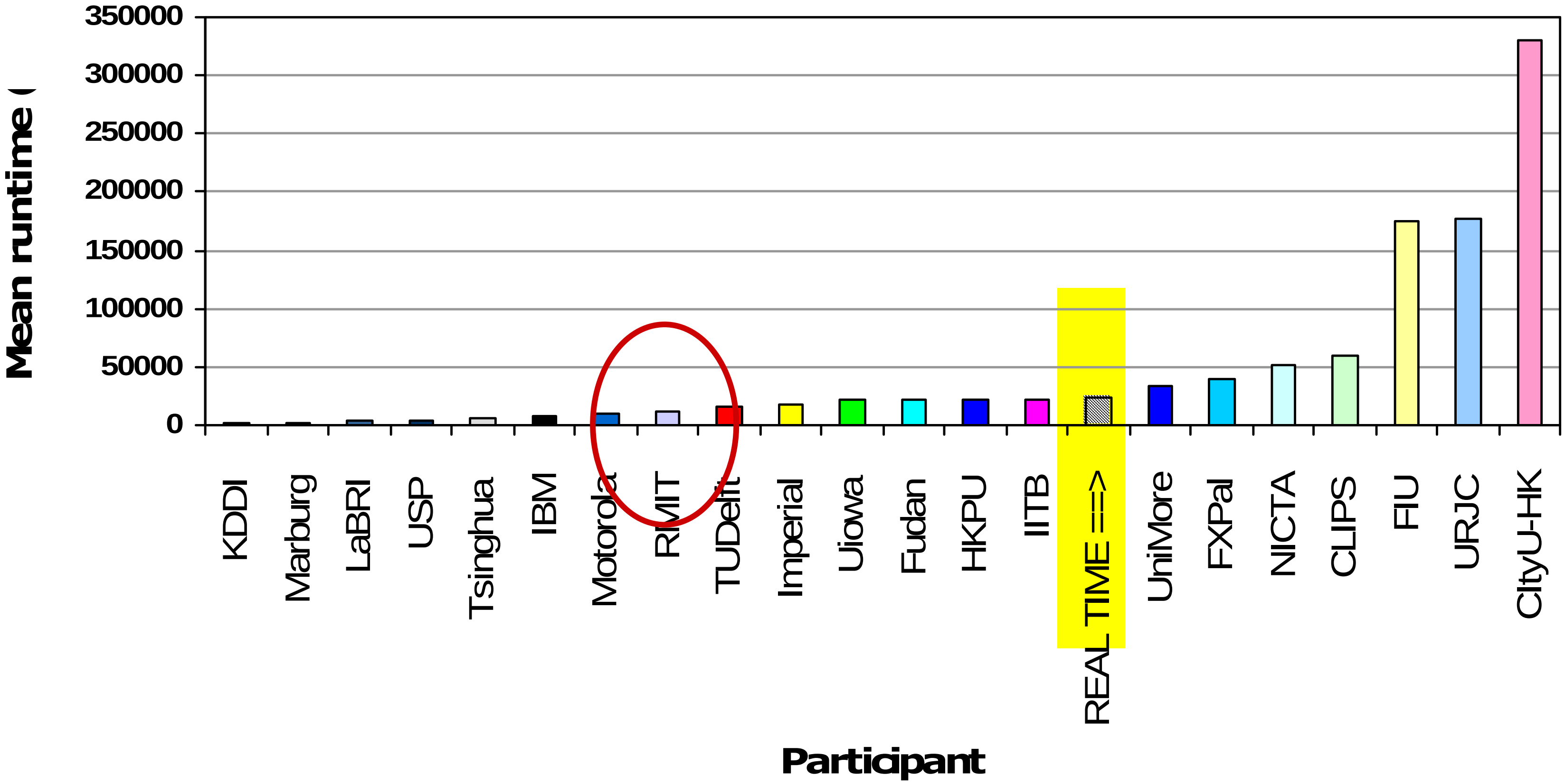
Cuts (zoomed again)



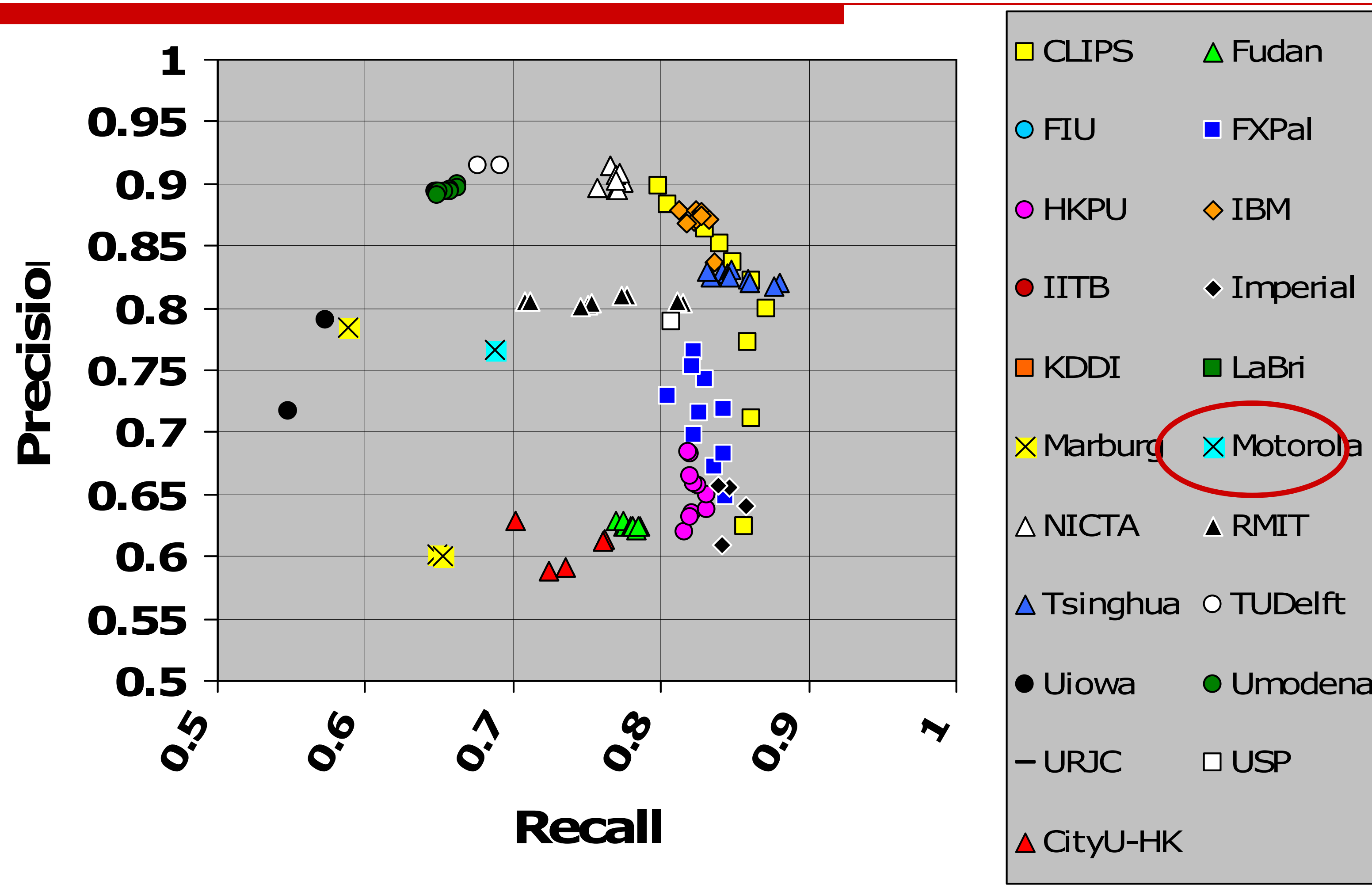
Gradual transitions (zoomed)



Mean runtime in seconds



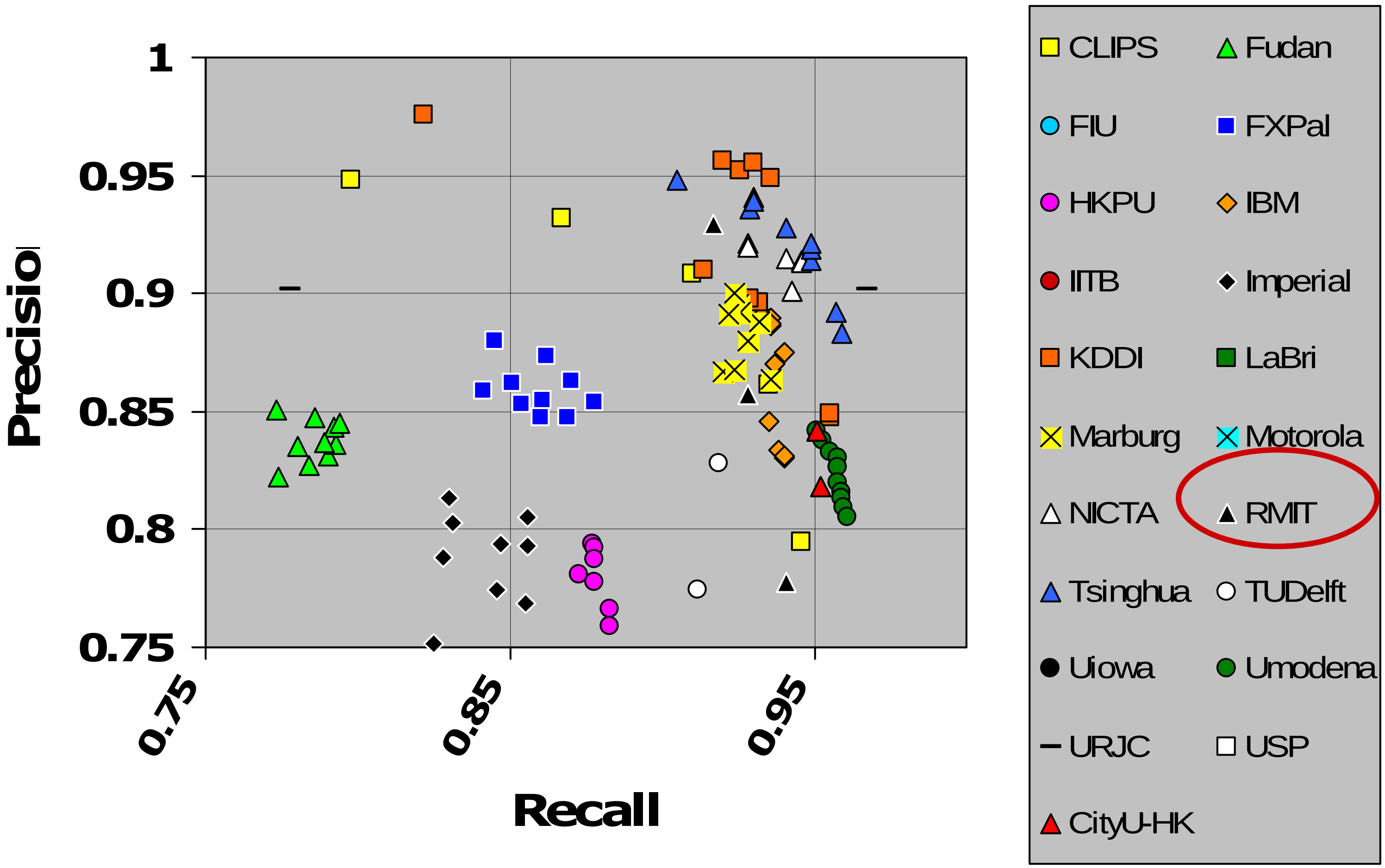
Gradual transitions: Frame-P & R (zoomed)



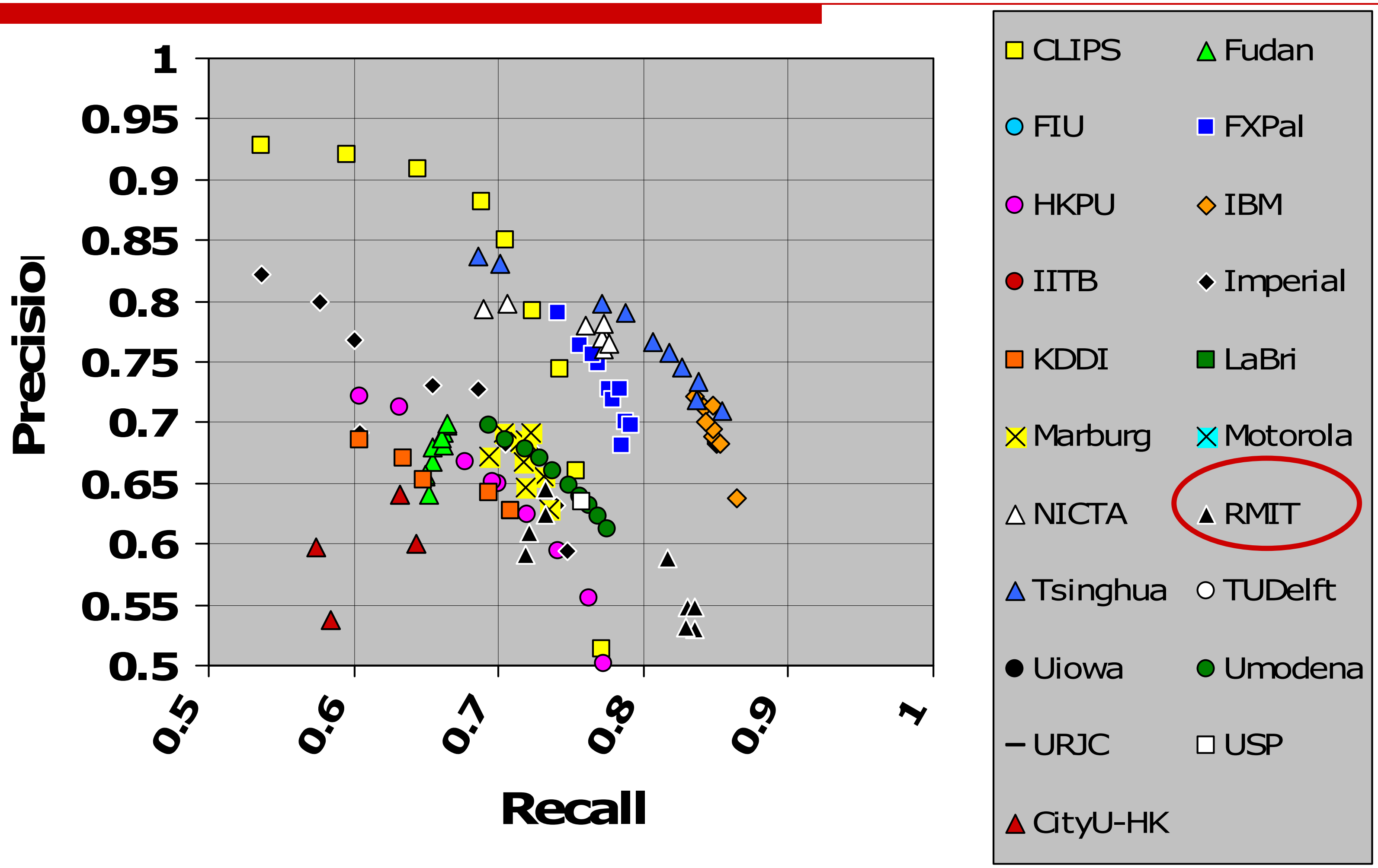
15. Technical University of Delft

- Approach
  - Represents video as spatio-temporal video data blocks and extracts patterns from these to indicate cuts and GTs;
- Performance
  - Efficient, expect to include camera motion information in future development;
- Results

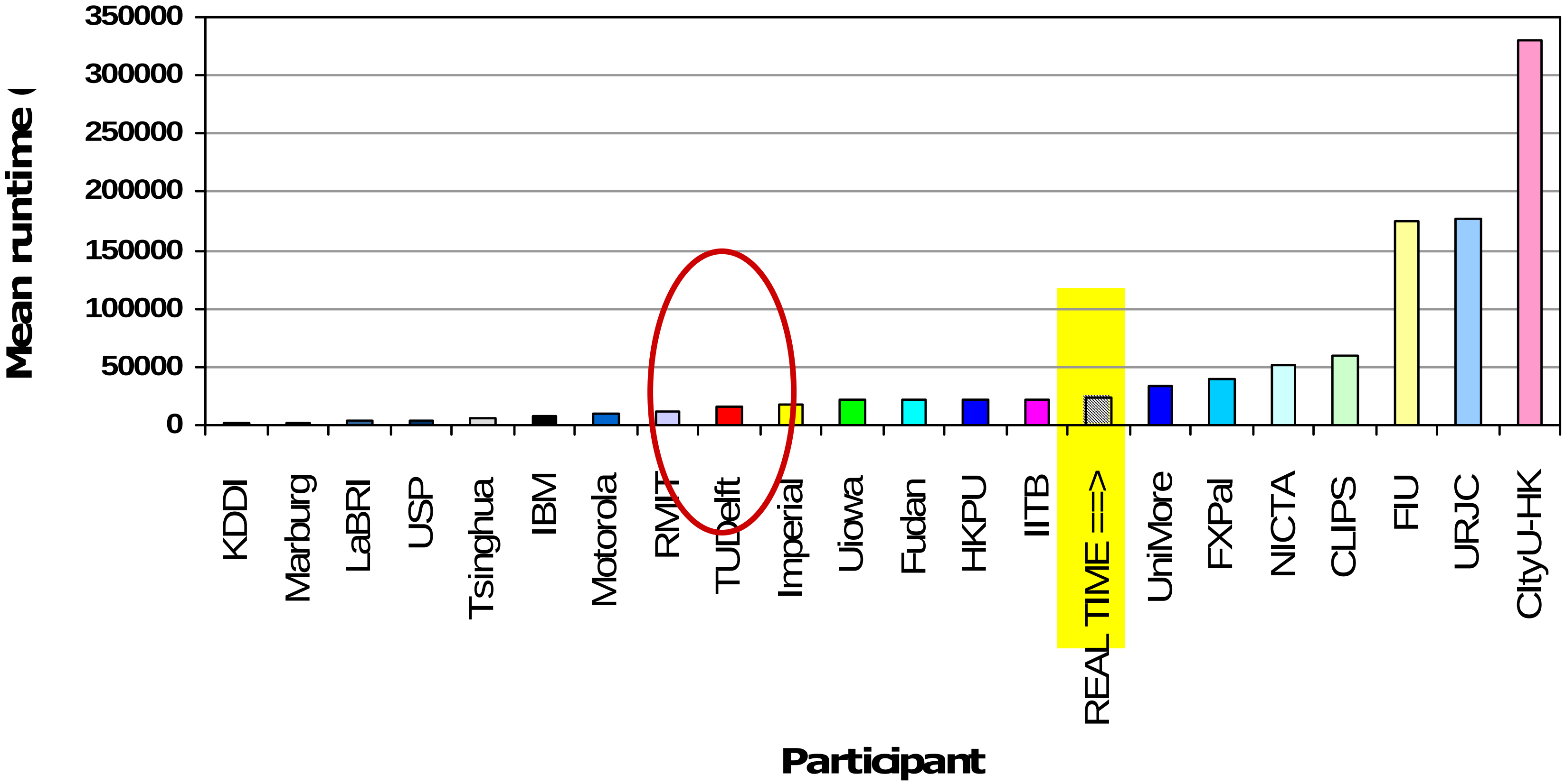
Cuts (zoomed again)



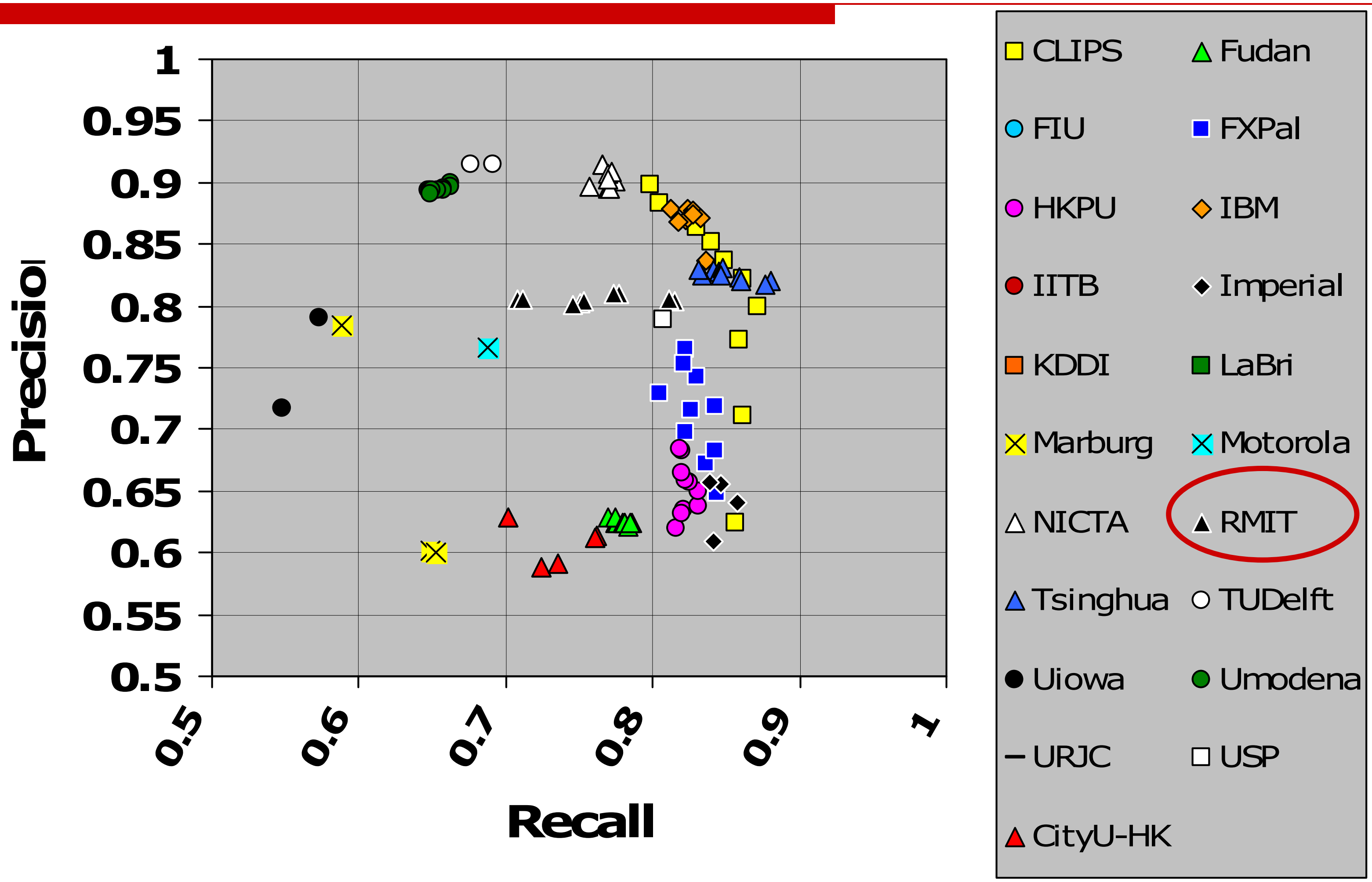
Gradual transitions (zoomed)



Mean runtime in seconds



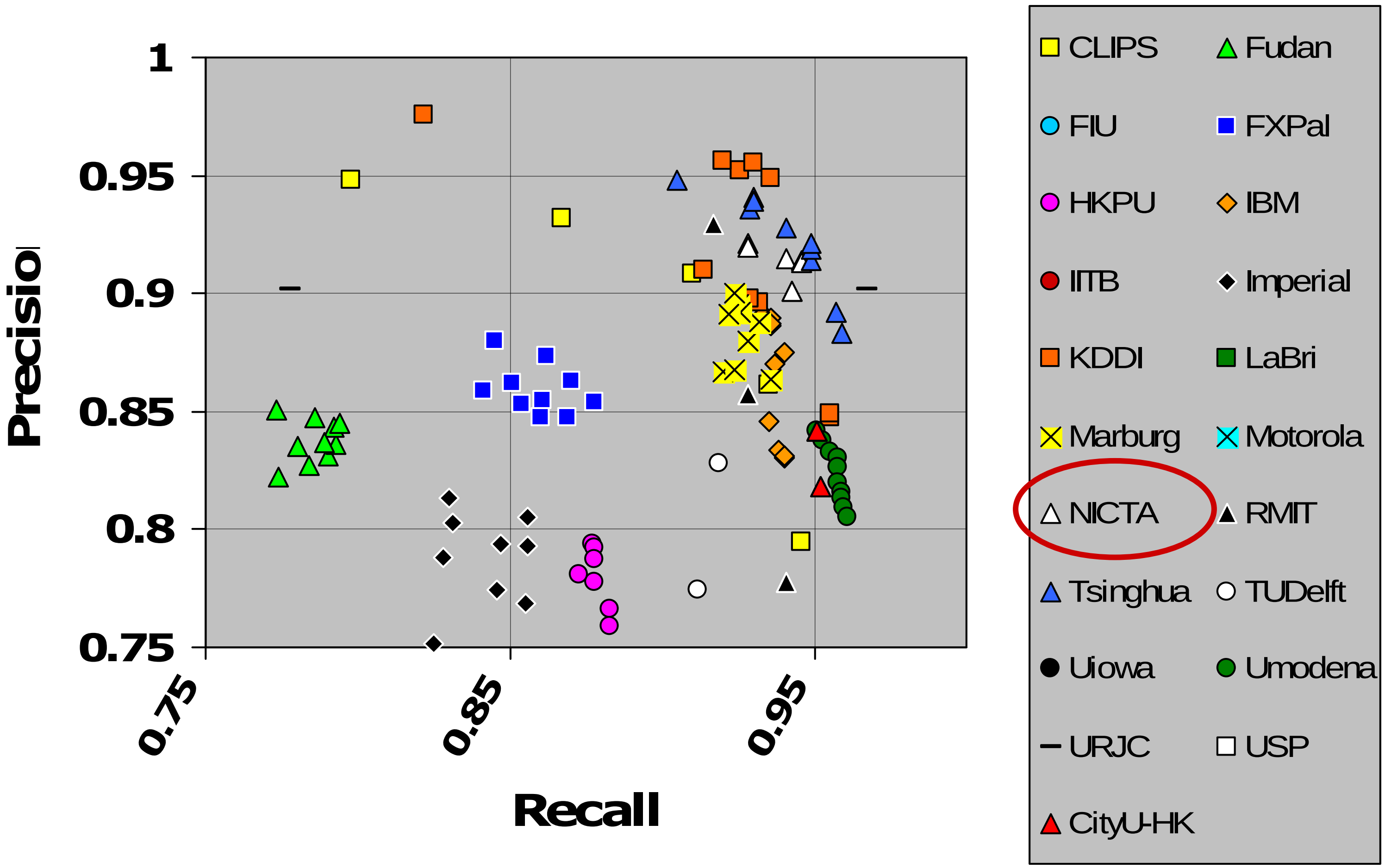
Gradual transitions: Frame-P & R (zoomed)



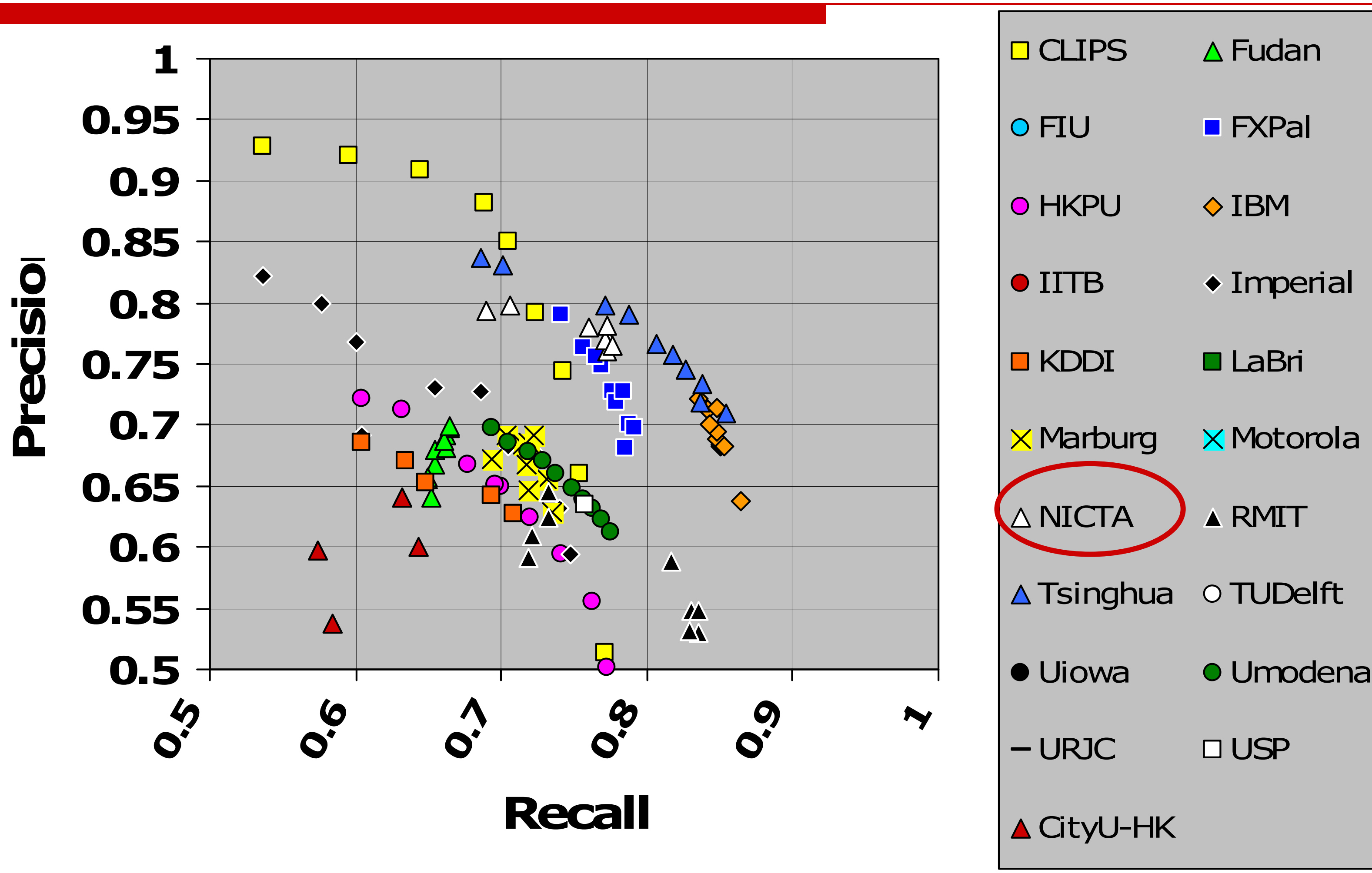
16. Tsinghua University

- Approach
  - Re-implement previous years very successful approaches which had evolved to a set of collaboration rules for various detectors;
  - Now a unified framework with SVMs combining fade-in/out detectors, GT detector and cut detectors, each developed in previous years;
- Features
  - Appears to be a mixture of different detectors;
- Performance
  - Despite individual detectors performing separately, very fast;
- Results

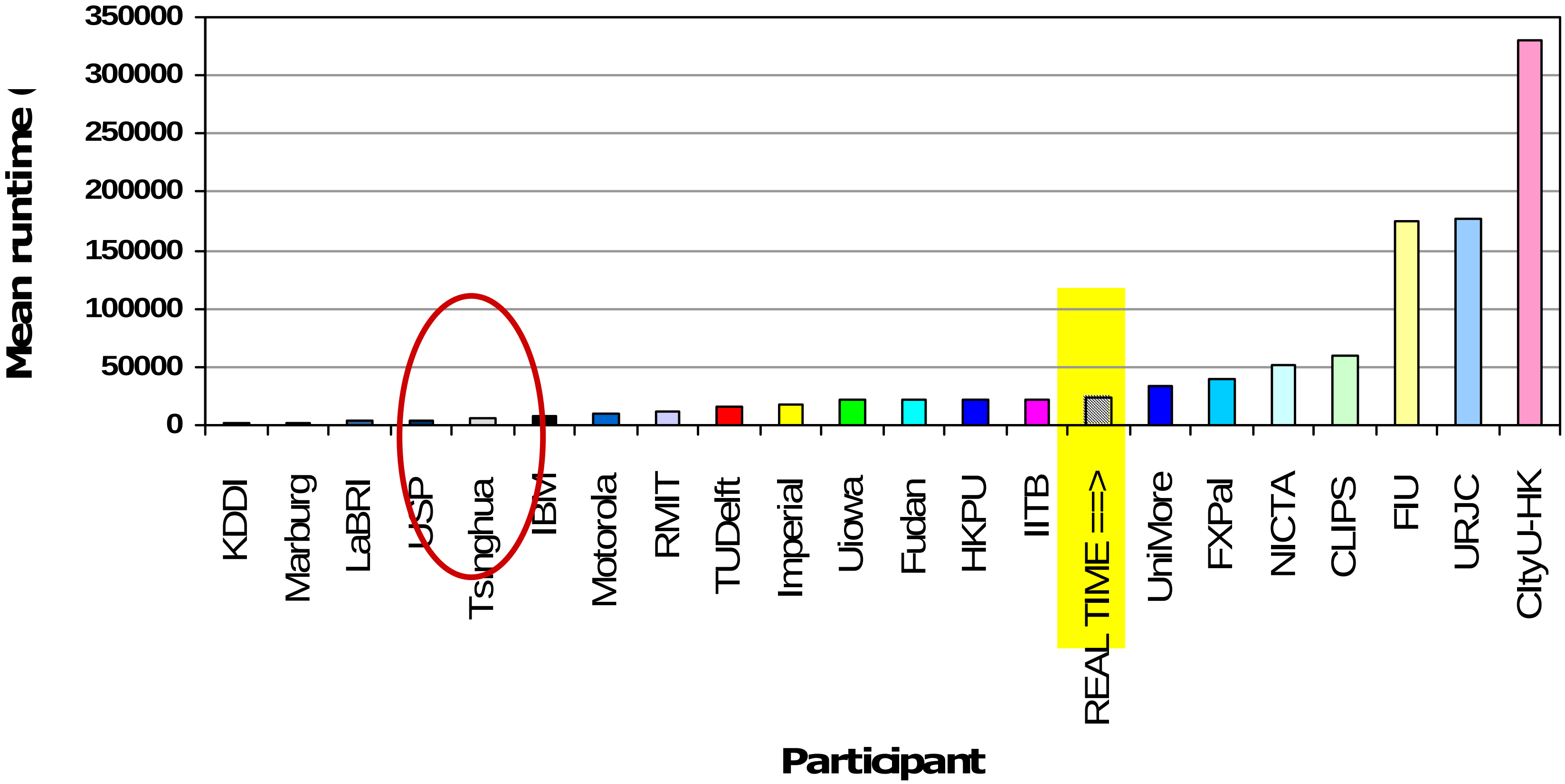
Cuts (zoomed again)



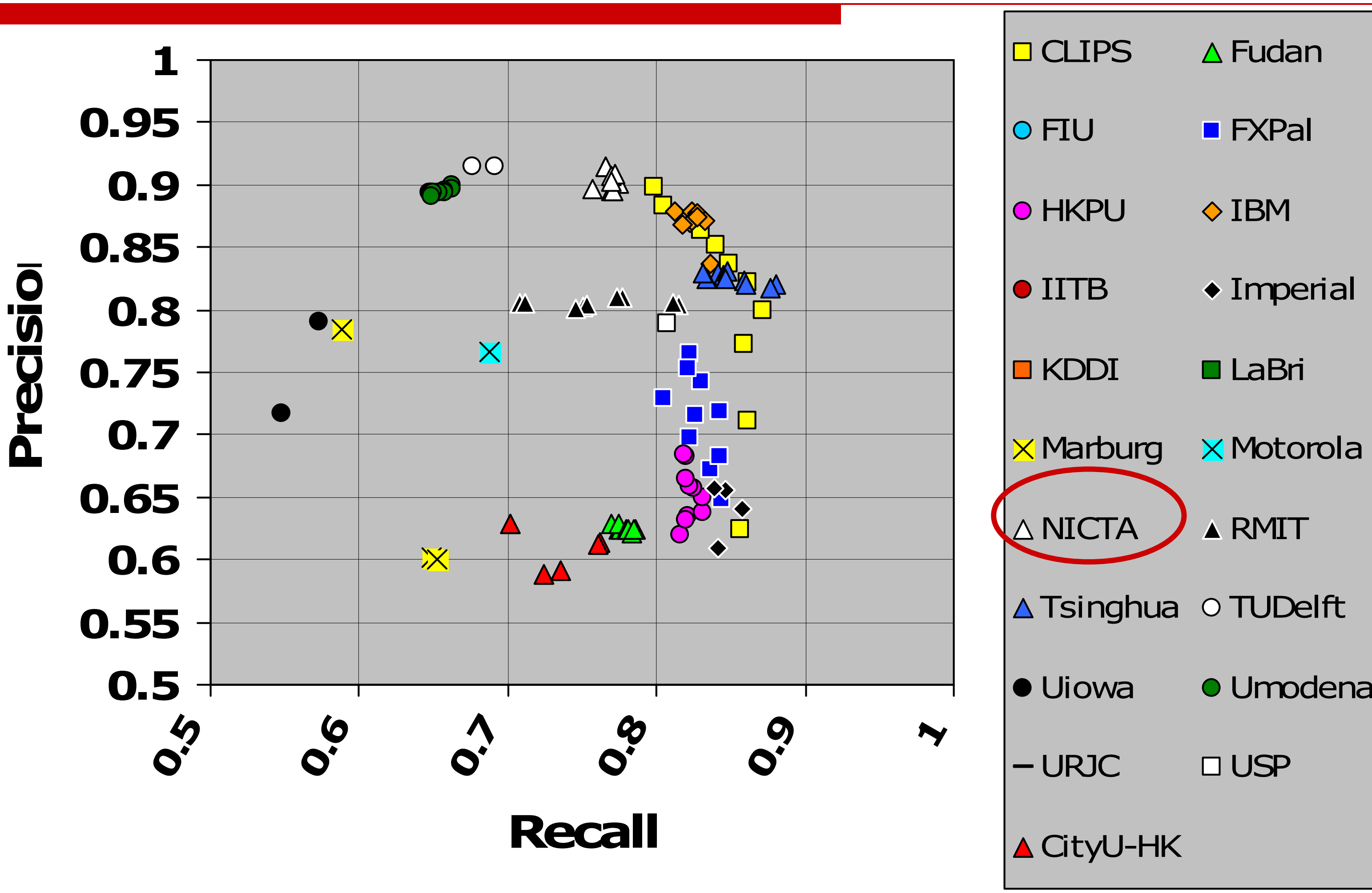
Gradual transitions (zoomed)



Mean runtime in seconds



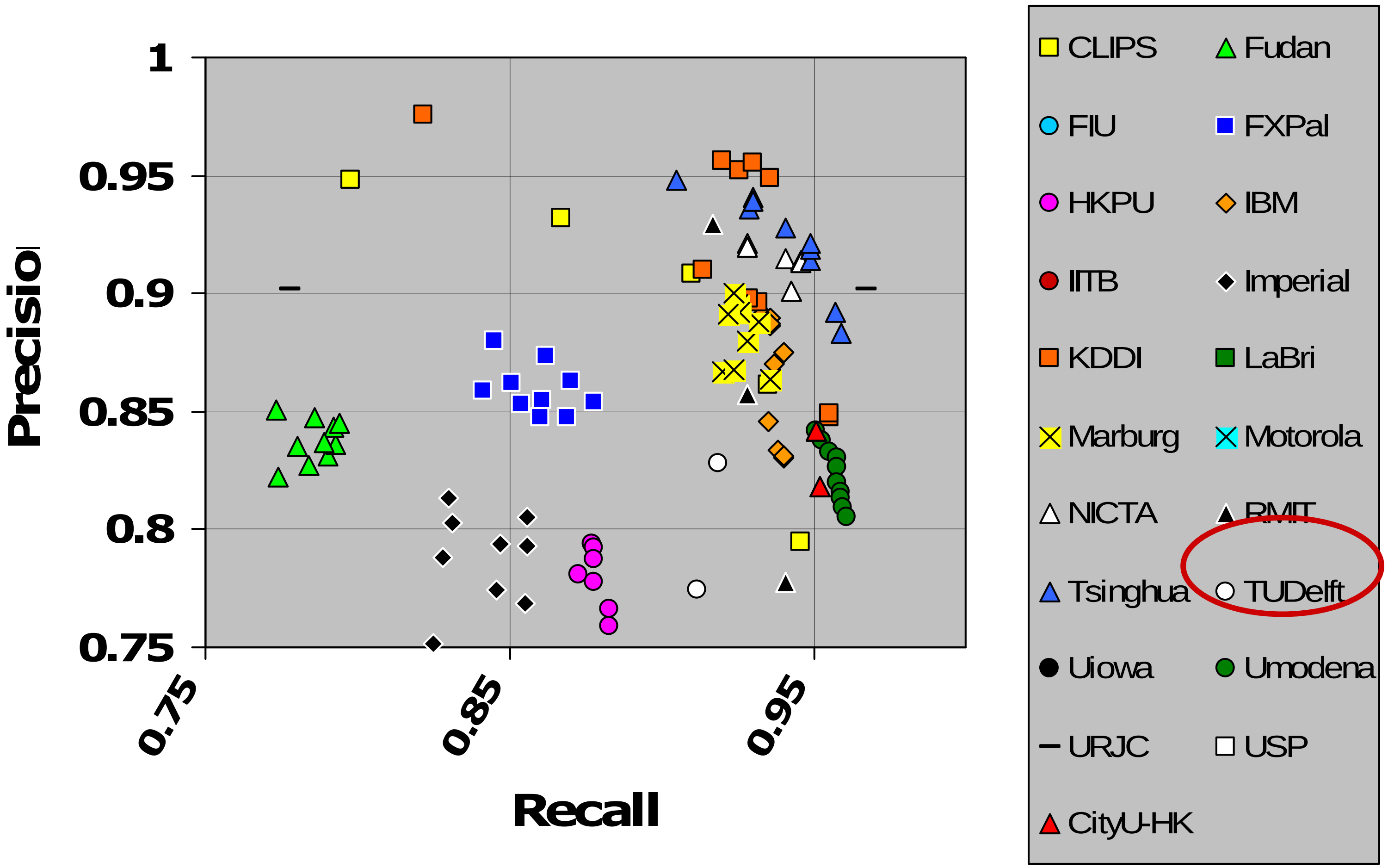
Gradual transitions: Frame-P & R (zoomed)



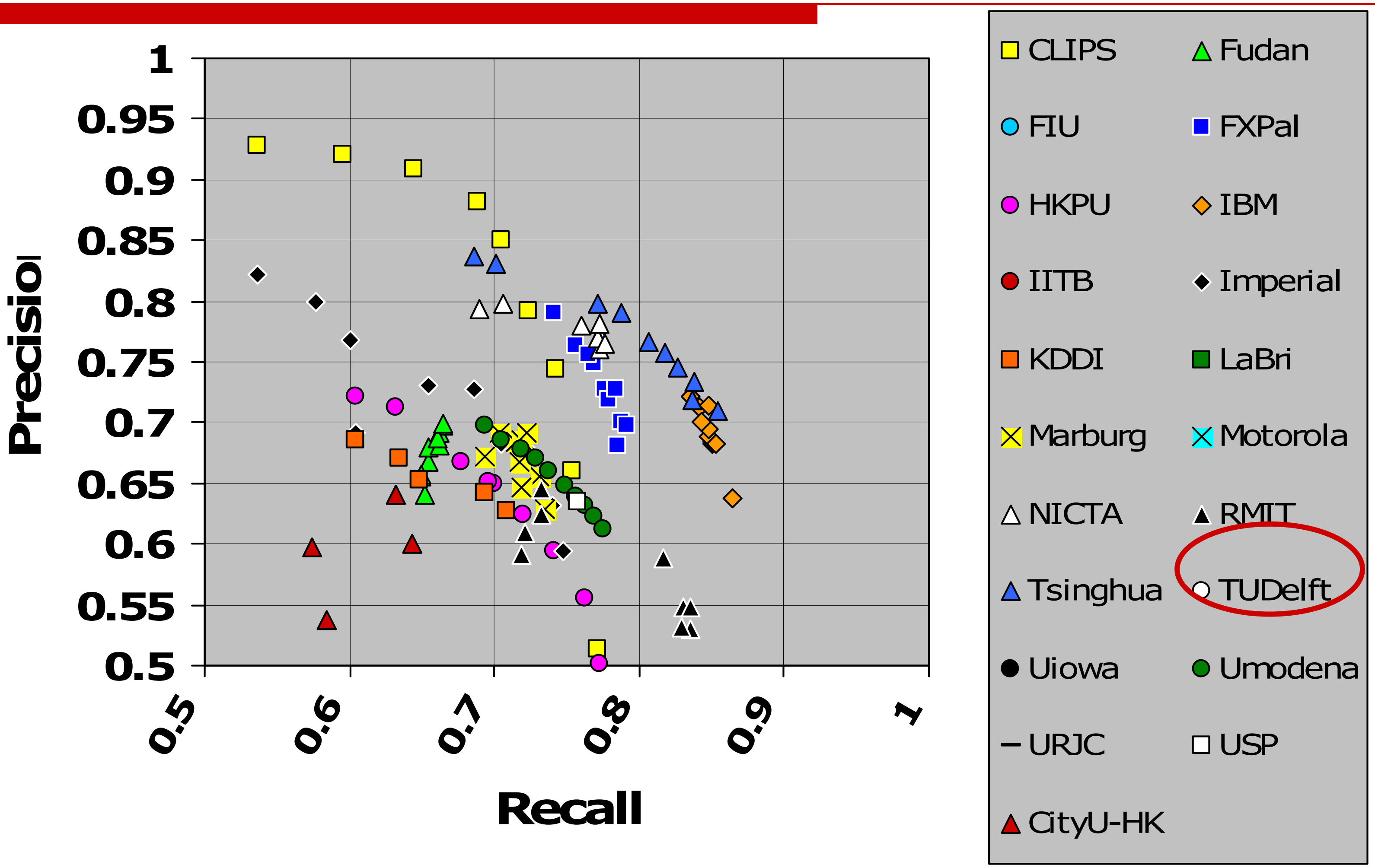
17. University of Central Florida/U. Modena

- Approach
  - Frame-frame distances computed based on pixels, and based on histograms;
  - Examined frame difference behaviours over time to see if it corresponds to a linear transformation;
- Features
  - Work carried out by U Modena;
- Performance
  - Could be speeded up but no optimisation;
- Results

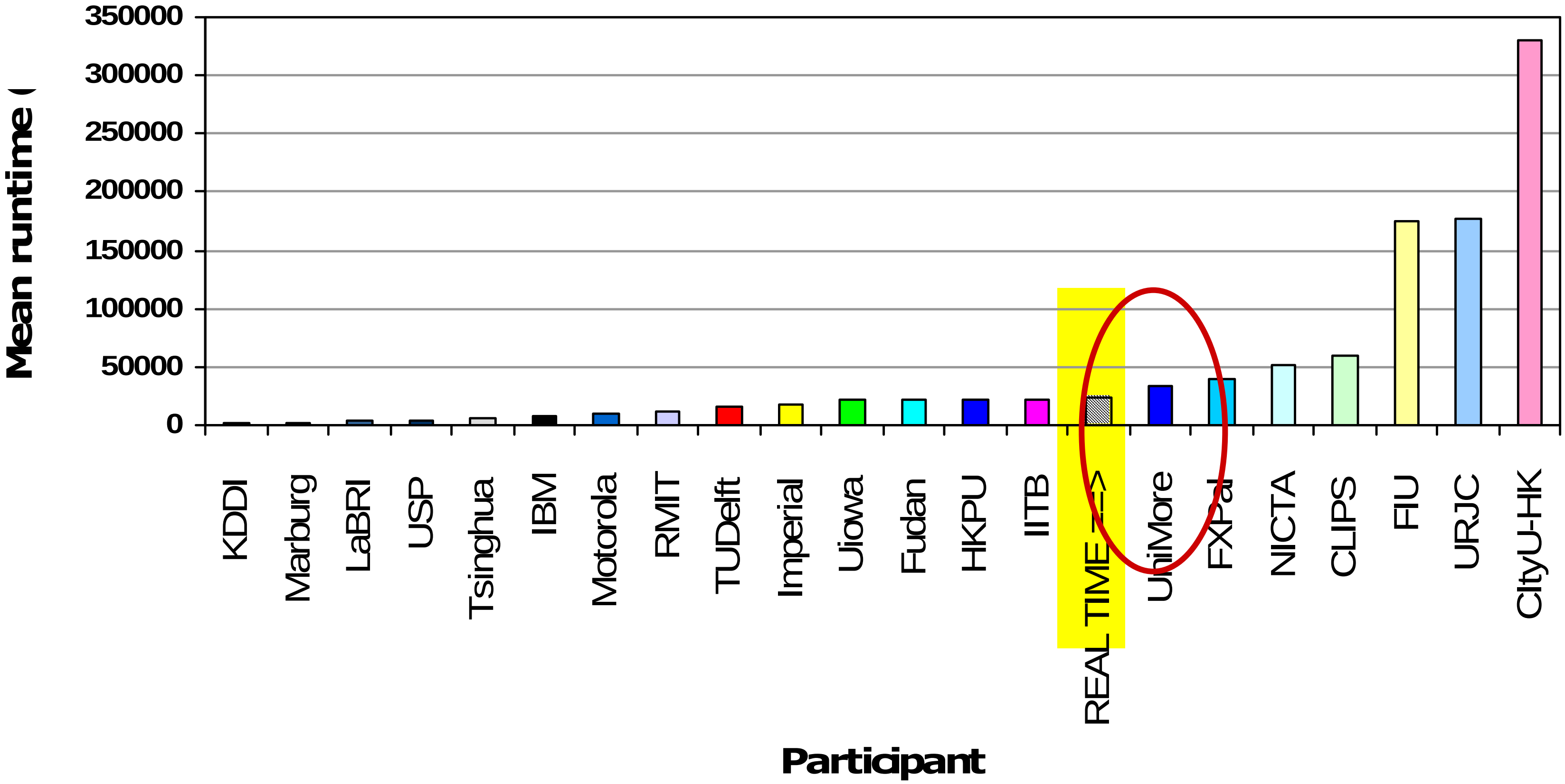
Cuts (zoomed again)



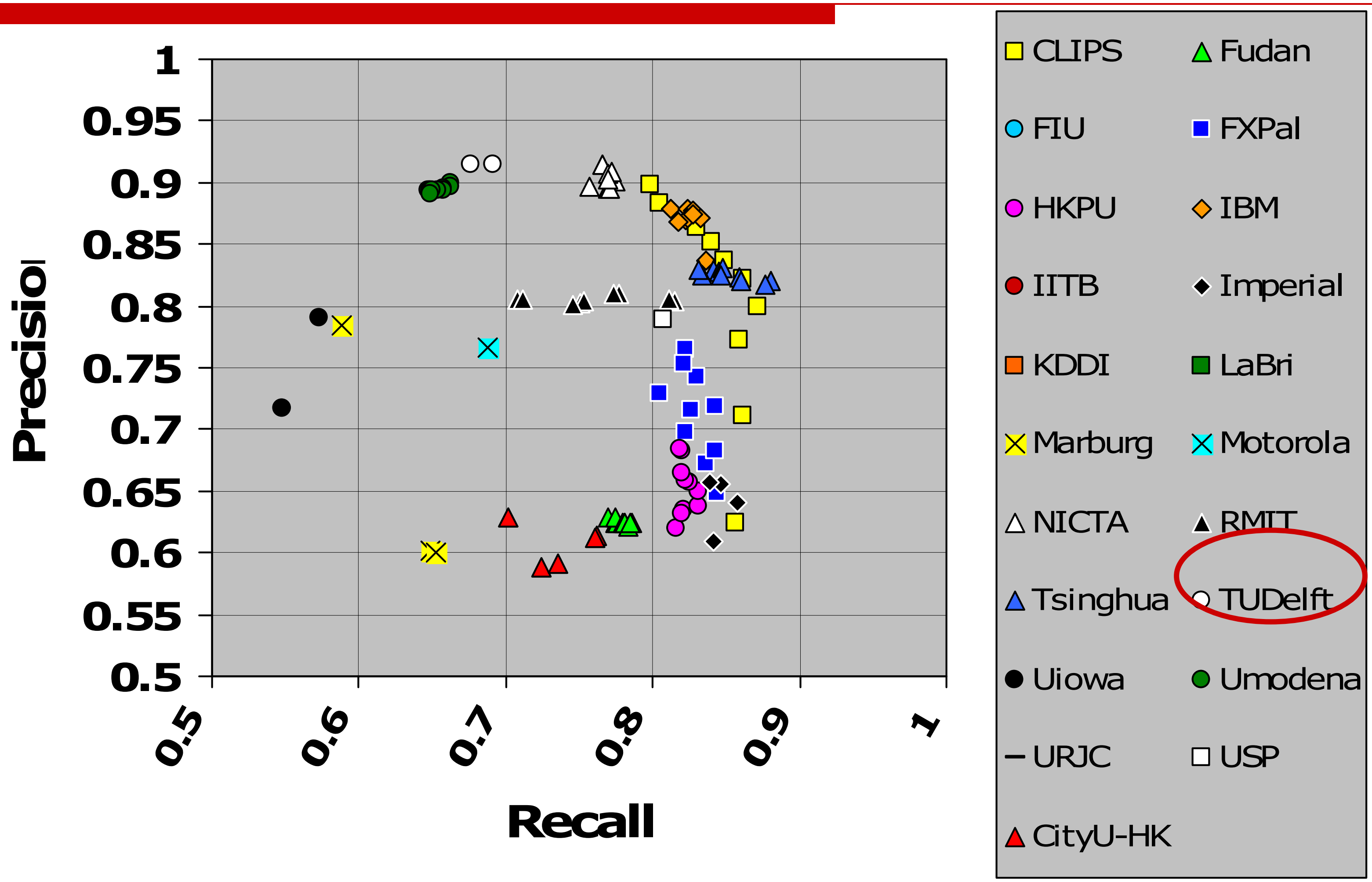
Gradual transitions (zoomed)



Mean runtime in seconds



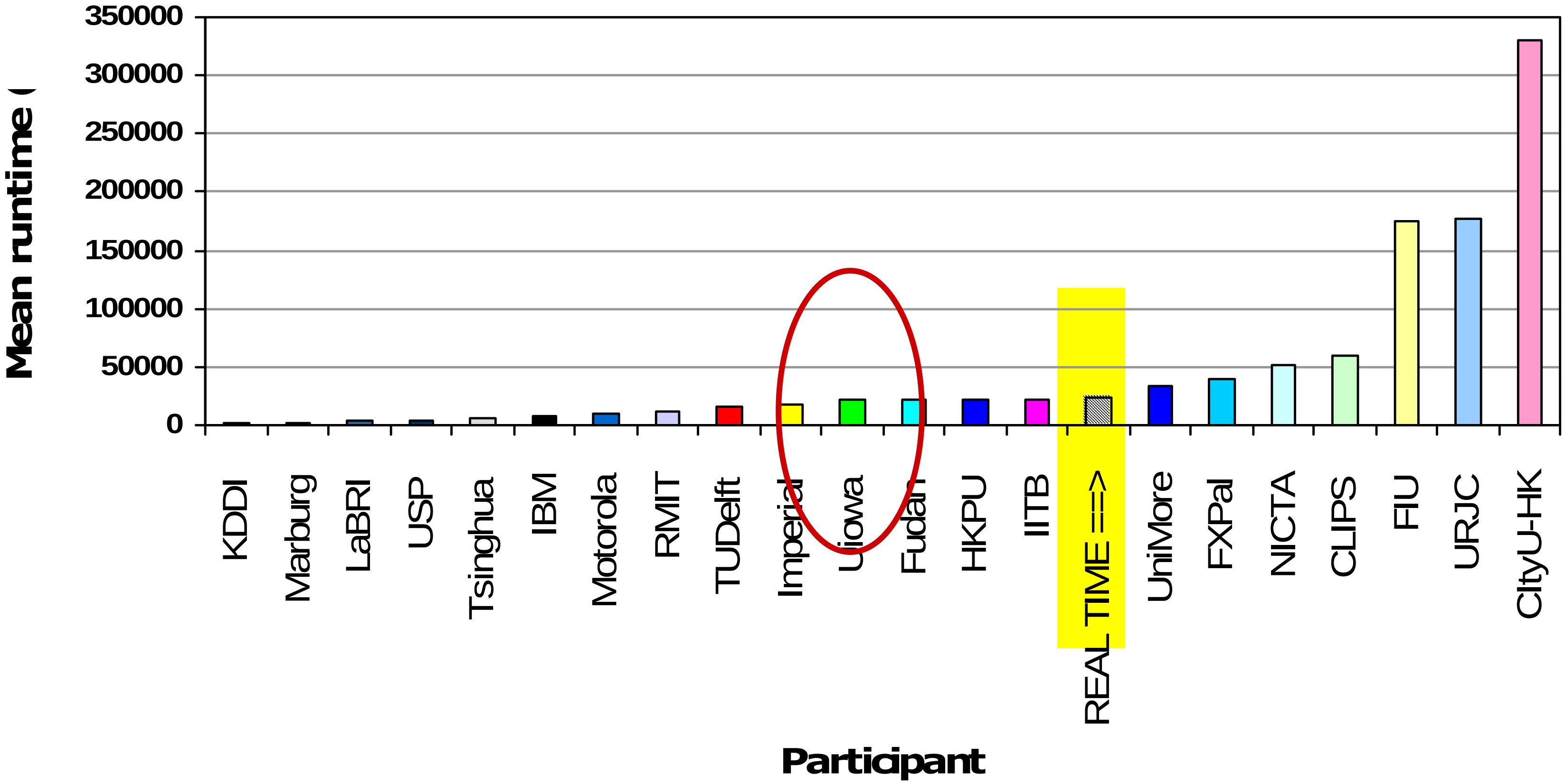
Gradual transitions: Frame-P & R (zoomed)



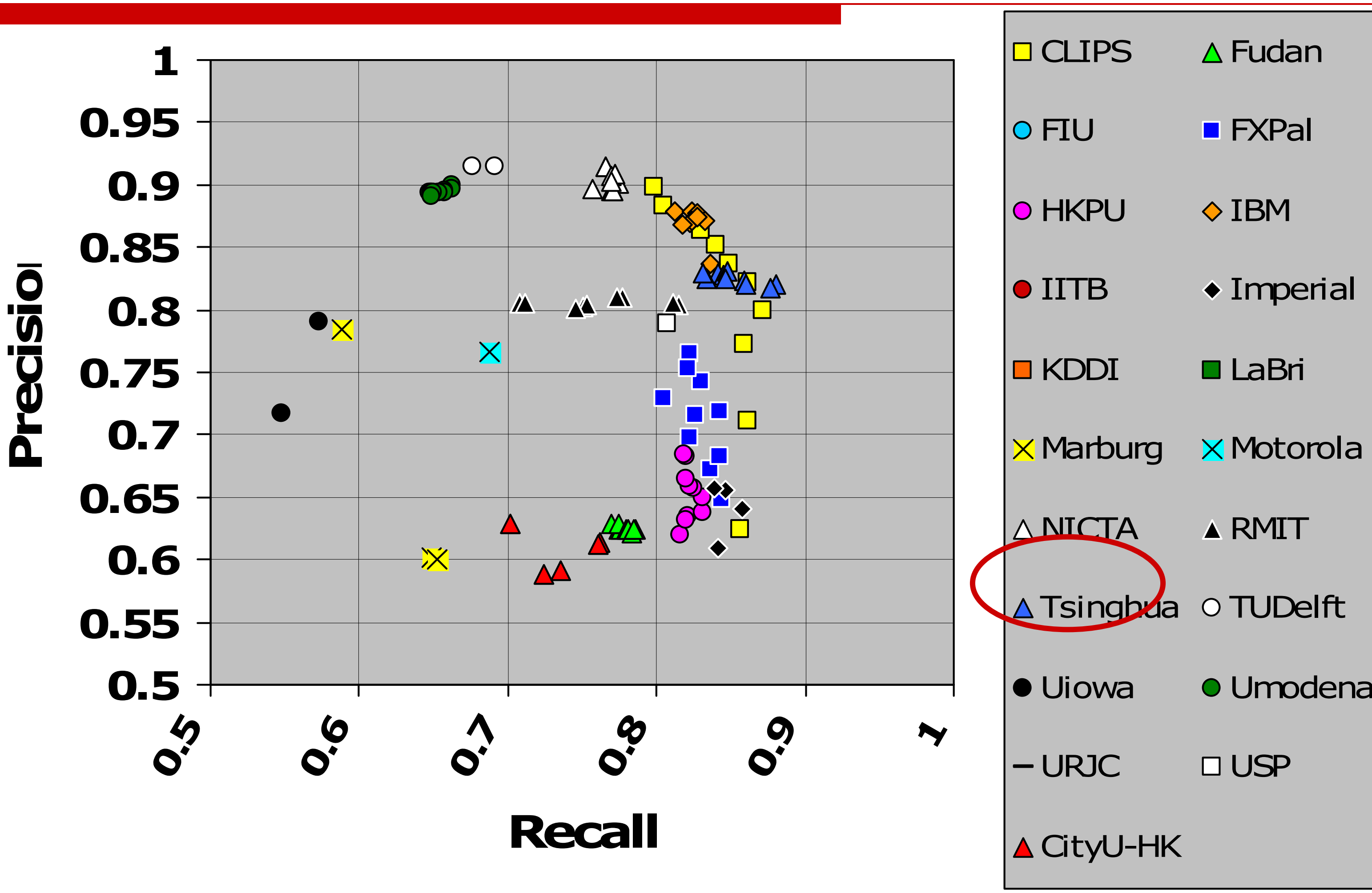
18. University of Iowa

- Approach
  - Builds upon previous years with a cut detection followed by GT detection;
  - Frame similarities based on colour histograms, on aggregated pixel distances and on edges;
- Performance
  - Still some issues of combining GT and cut logic detection, not appearing in zoomed areas of graphs;
- Results

Mean runtime in seconds



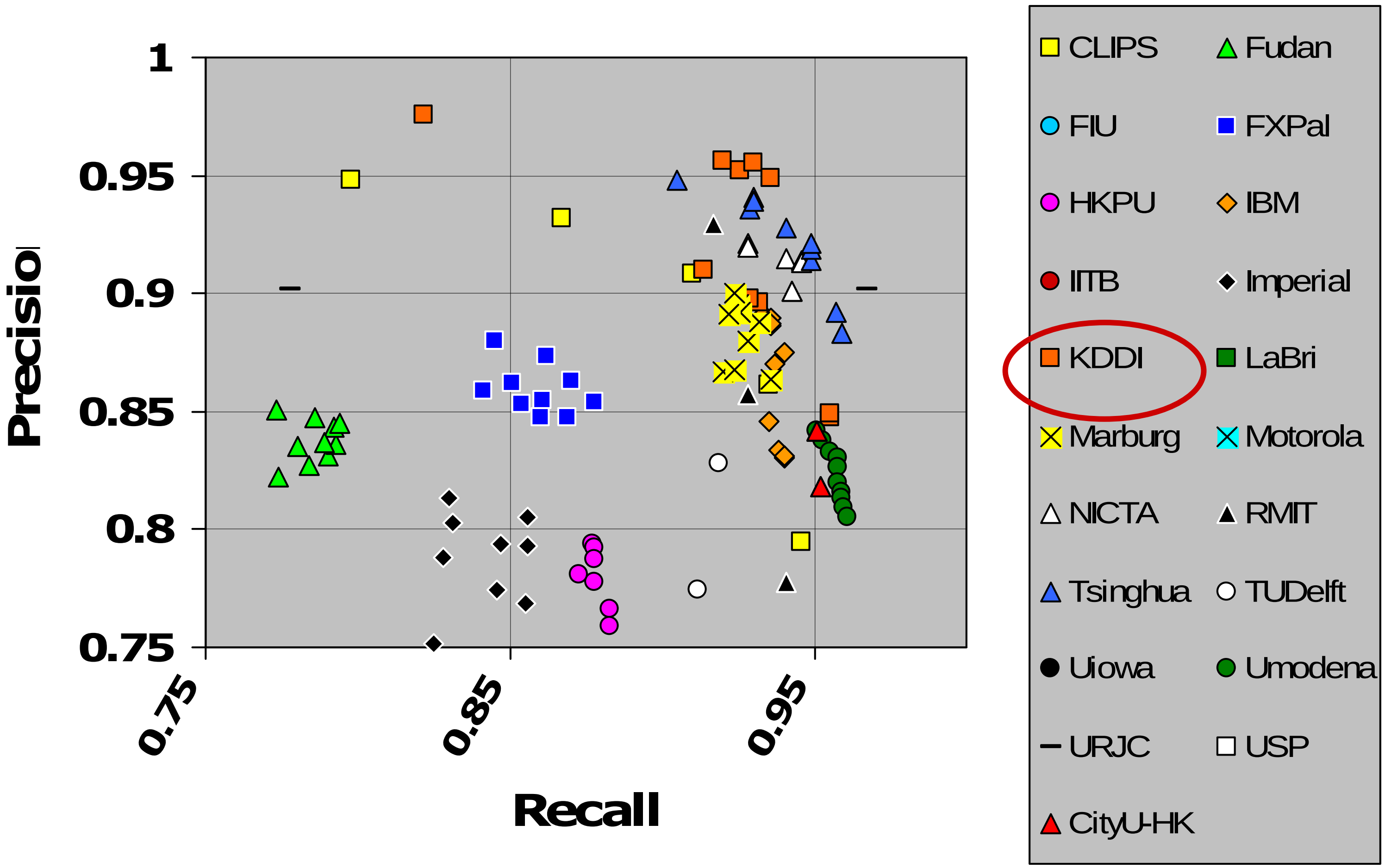
Gradual transitions: Frame-P & R (zoomed)



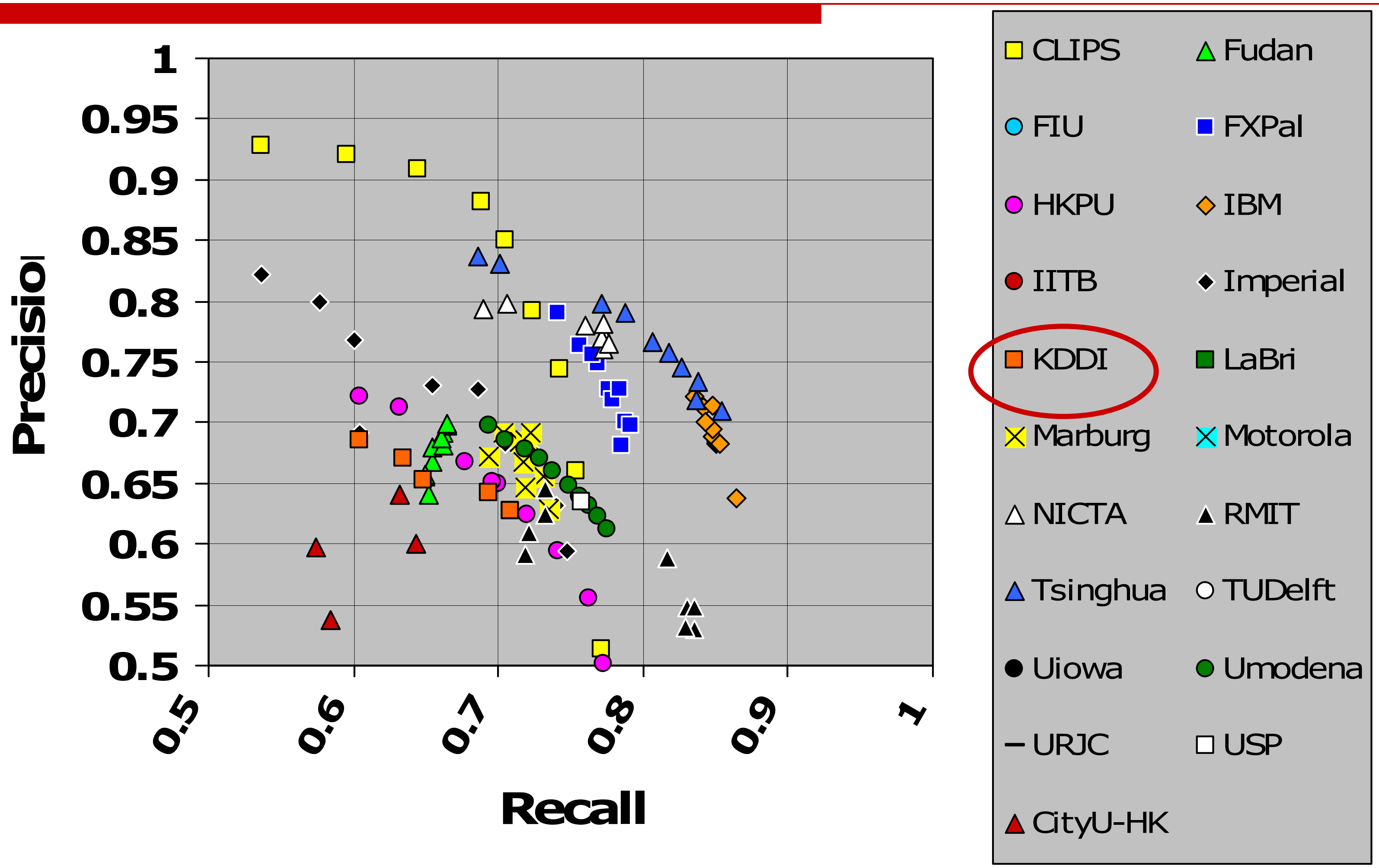
19. University of Marburg

- Approach
  - Frame similarities measured by motion-compensated pixel differences and histogram differences for several frame distances;
  - An unsupervised ensemble of classifiers is then used.
- Features
  - SVM classifiers trained on 2004 data;
- Performance
  - Surprisingly efficient and good performance;
- Results

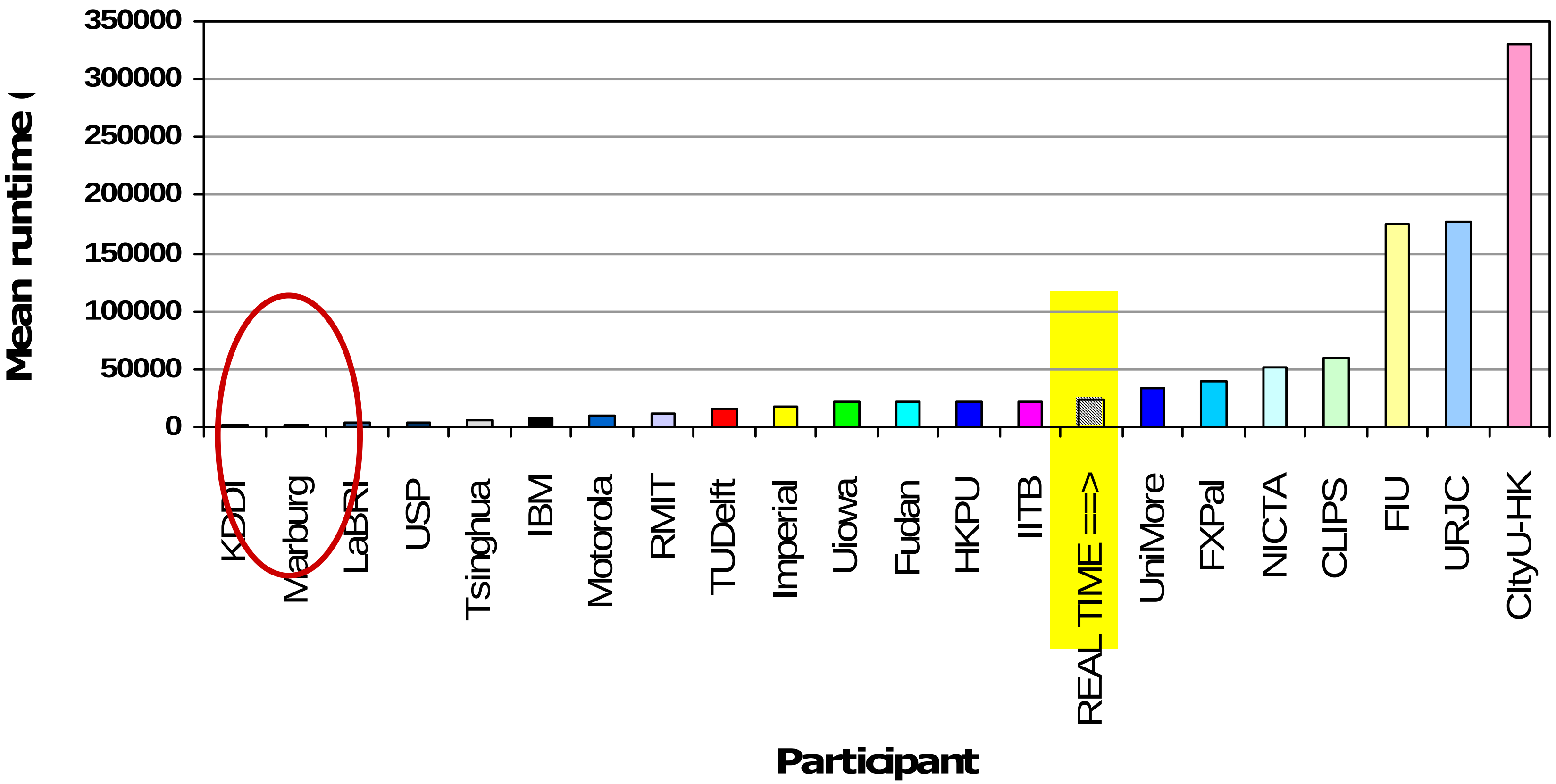
Cuts (zoomed again)



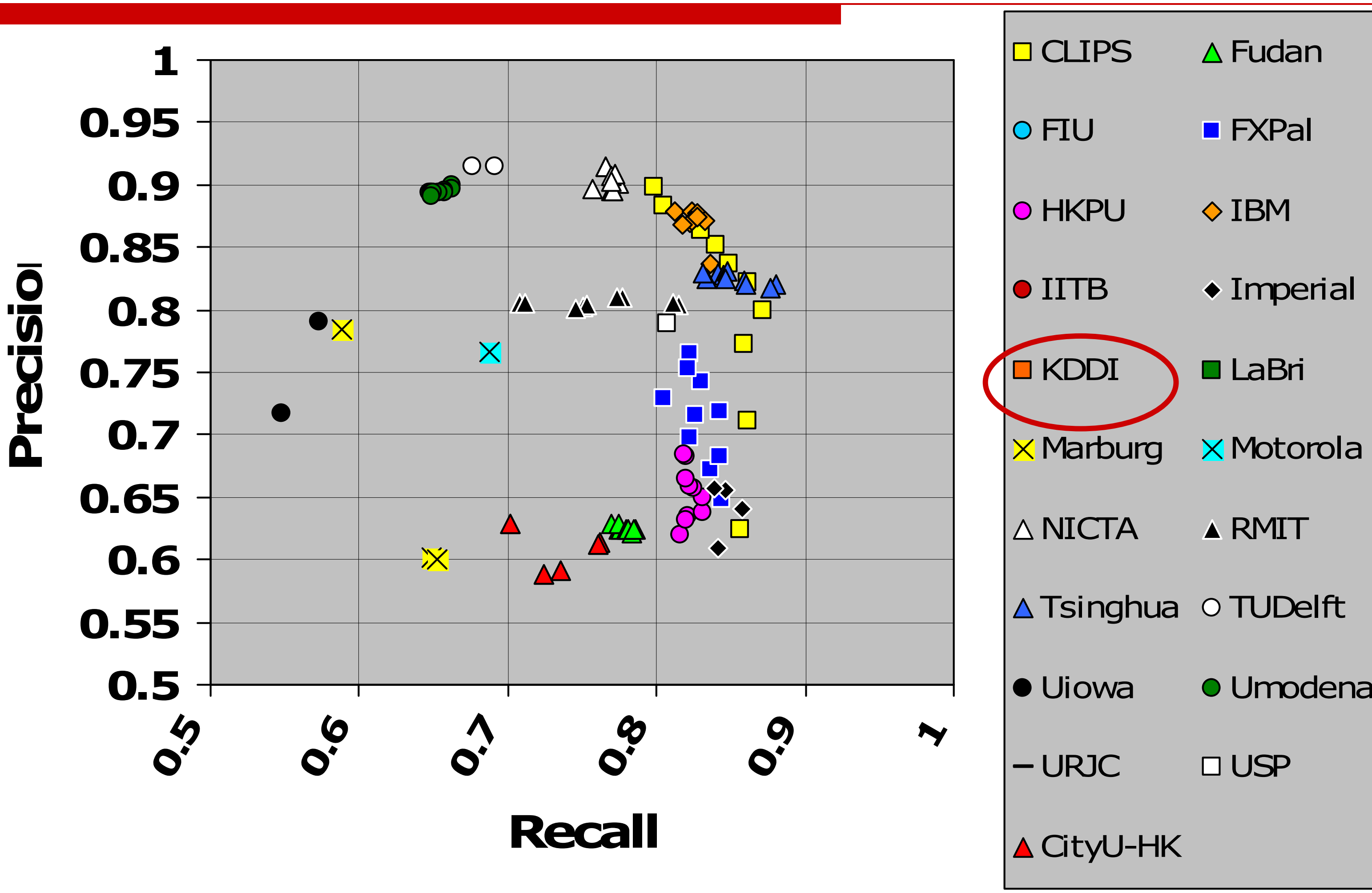
Gradual transitions (zoomed)



Mean runtime in seconds



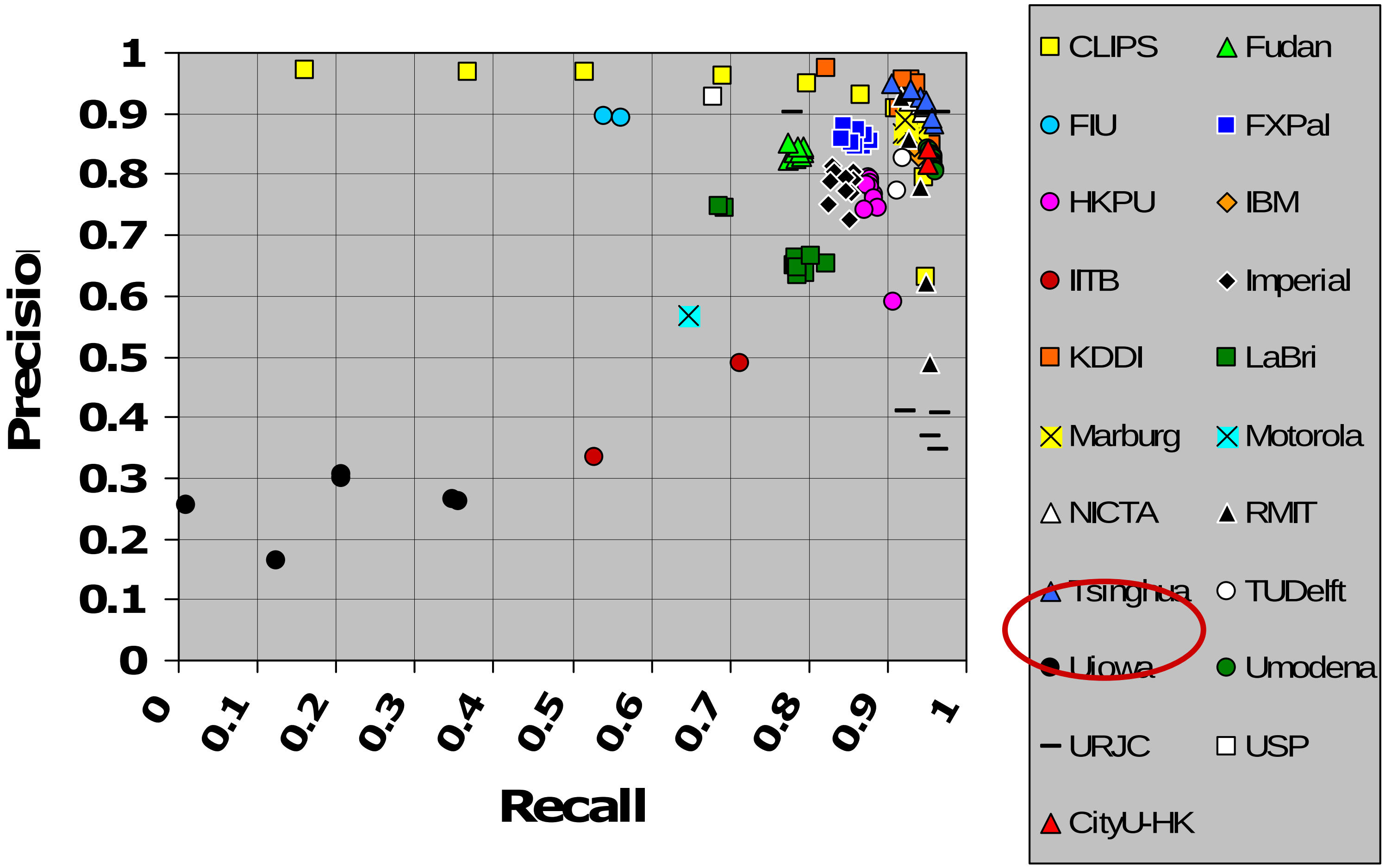
Gradual transitions: Frame-P & R (zoomed)



20. University Rey Juan Carlos

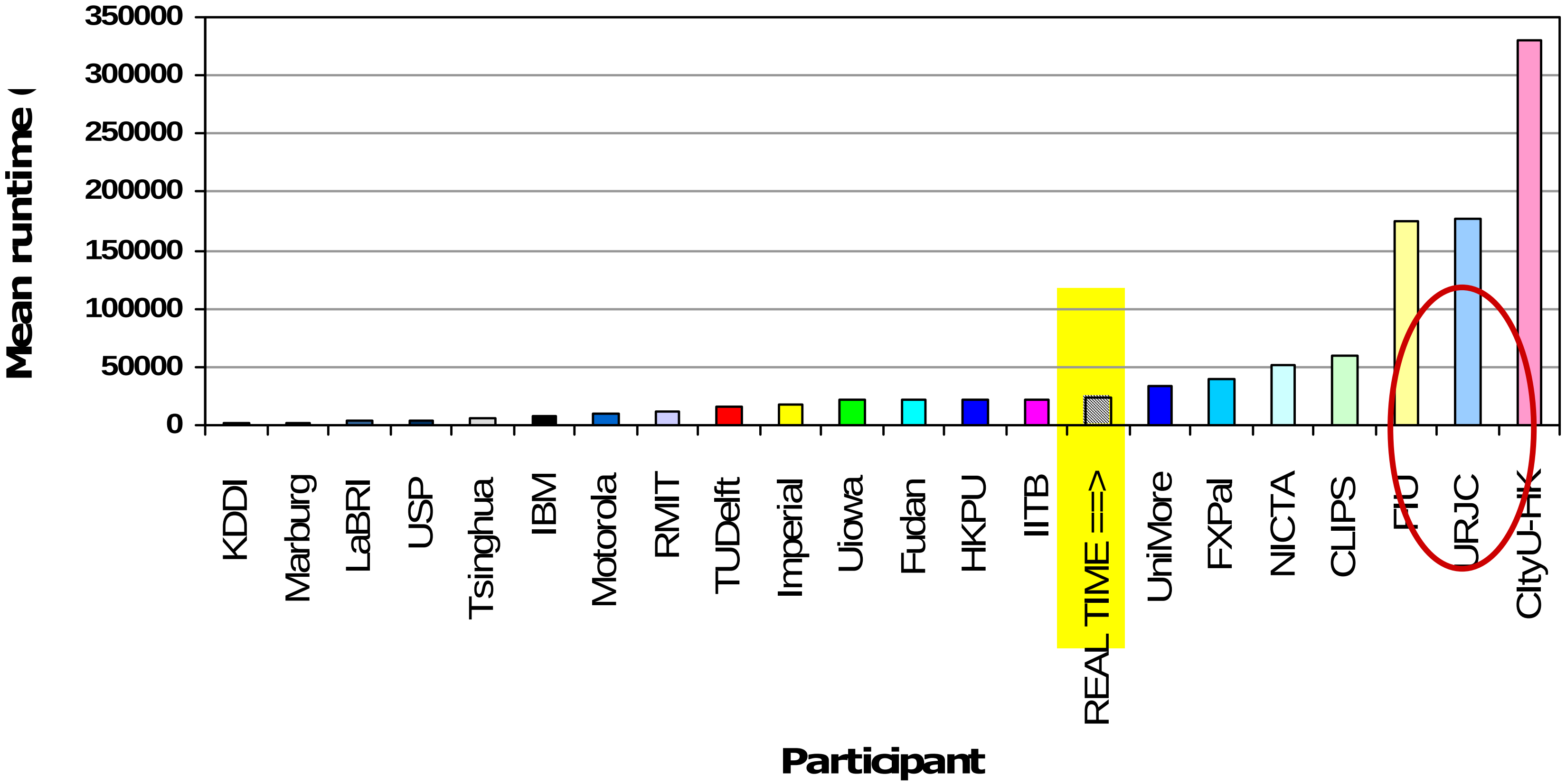
- Approach
  - Concentrated on cut detection by shape and by a combination of shape and colour features;
  - Shape used Zernike moments, colour used histograms from last year;
  - Combination methods used various logical combinations
- Performance
  - Did well on precision for cuts, not in zoomed areas otherwise;
- Results

Cuts



Precision
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
CLIPS
Fudan
FIU
FXPal
HKPU
IBM
IITB
Imperial
KDDI
LaBri
Marburg
Motorola
NICTA
RMIT
Tsinghua
TUDelft
UIowa
Umodena
URJC
USP
CityU-HK

Mean runtime in seconds

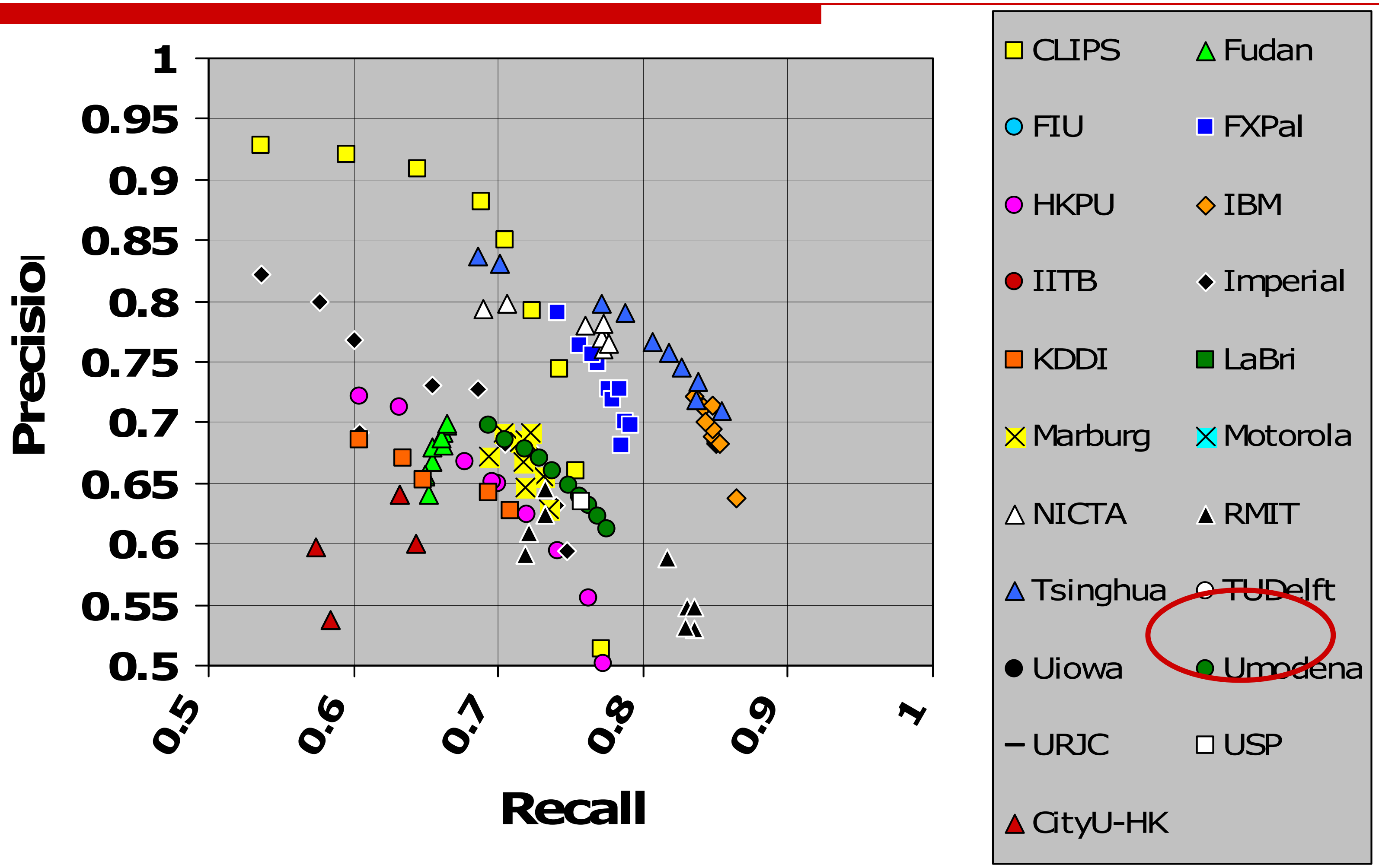


21. Universidade São Paulo

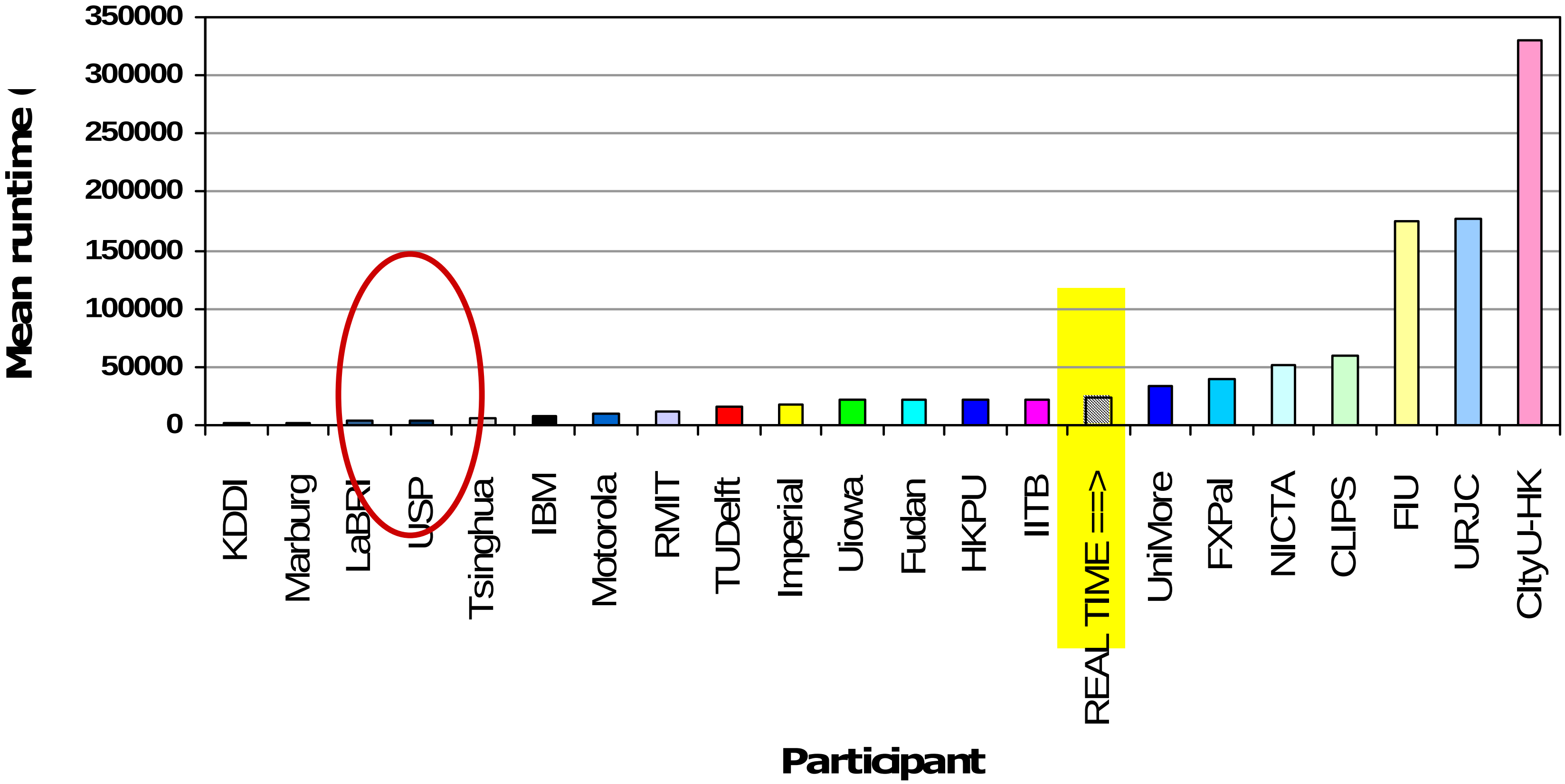
- Approach
  - No paper submitted - again - so we don’t know

- Results
  - Appears to be fast and appearing in the zoomed areas of the graphs;

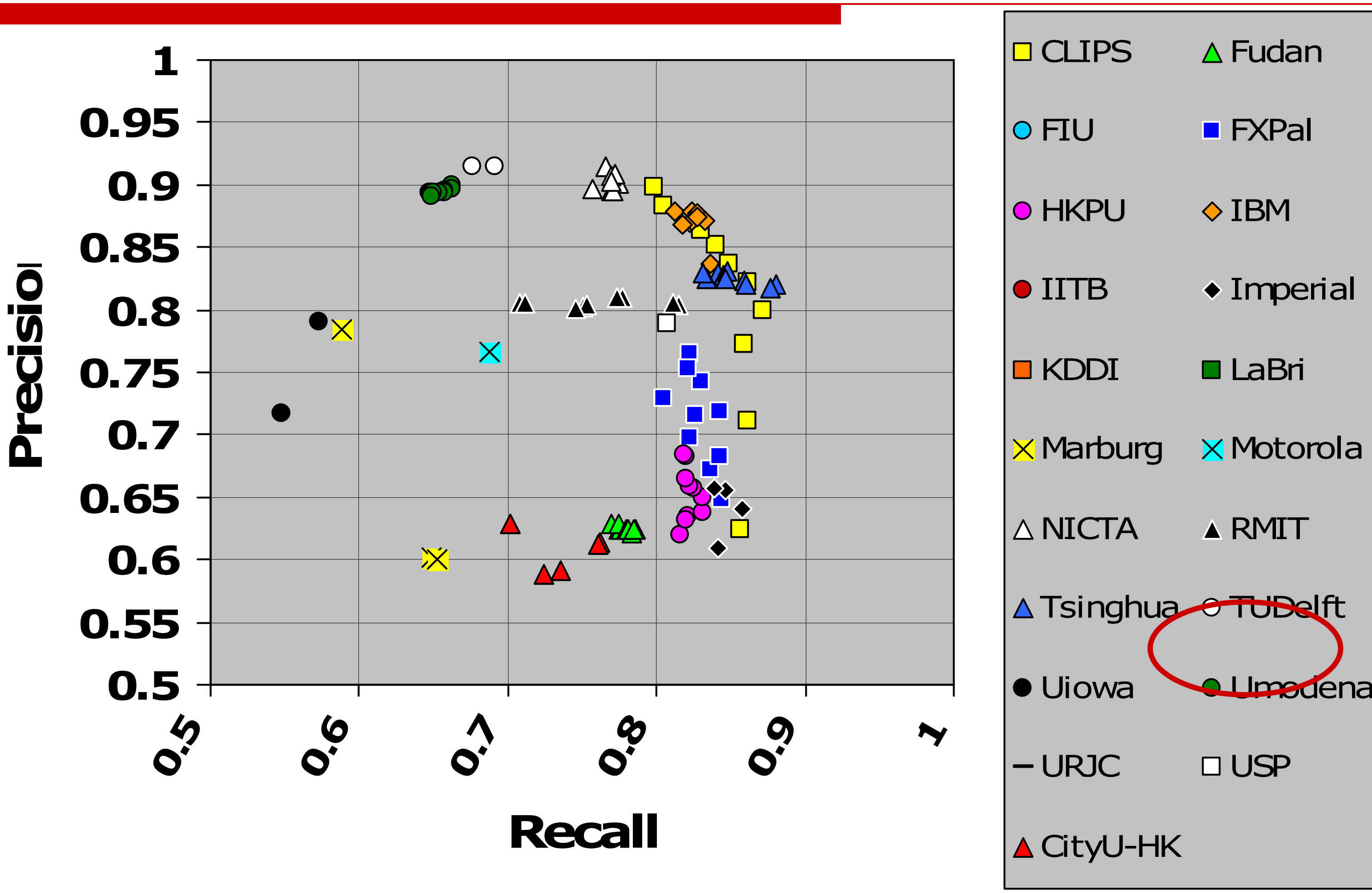
Gradual transitions (zoomed)



Mean runtime in seconds



Gradual transitions: Frame-P & R (zoomed)



Observations

- Last year we said:
  - Strong interest;
    - … this remains true … more in SBD than in search in TRECVid2005 … regulars, devotees, and new participants;
  - Novel approaches continue to emerge;
    - … absolutely true still … new things still being tried;
  - Adding computation cost was a good idea;
    - … and it remains an interesting & important criterion;
  - Lots of data available to do a more comprehensive comparative analysis;
    - … though nobody has done this yet;

Conclusions

- What did we learn this year ?
  - More new and faster methods, and data didn't throw us any surprises, though maybe its quite similar to 2004/3 and NASA data didn't pollute it enough ?
- Some people ask why bother … isn't SBD a solved problem ?
  - Hard to argue against this when we can show excellent accuracy in a fraction of real-time for cuts - for GTs do we need better performance
  - Yet new approaches emerge each year, its very economical to run the task, and teams can break into video manipulation;
  - More groups do SBD than all search tasks combined !