TRECVID-2006: 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 (26)

1. AIIA Laboratory	**Greece**	**14. IIT / NCSR Demokritis**	**Greece**
2. AT&T Laboratories	**USA**	**15. KDDI / Tokushima U. / ISM / NII**	**Japan**
3. Chinese Academy of Sciences / JDL	**China**	**16. ETIS**	**Greece**
4. City University of Hong Kong	**China**	**17. Motorola Research Lab.**	**USA**
5. CLIPS-IMAG, LSR-IMAG	**France**	**18. RMIT University**	**Australia**
6. COST292	**EU**	**19. Tokyo Institute of Technology**	**Japan**
7. Curtin University	**Australia**	**20. Tsinghua University**	**China**
8. Dokuz Eylol	**Turkey**	**21. University of Marburg**	**Germany**
9. Florida International University	**USA**	**22. University of Modena Reggio**	**Italy**
10. FX Palo Alto Laboratory	**USA**	**23. Carleton University (Ottawa)**	**Canada**
11. Helsinki University of Technology	**Finland**	**24. University of Sao Paulo (USP)**	**Brazil**
12. Huazhong U. of Science & Tech.	**China**	**25. University Rey Juan Carlos**	**Spain**
13. Indian Institute of Tecnology, Bombay	**India**	**26. Zhejiang University**	**China**

2005 had 21 groups, of whom 9 appear again in 2006

Shot boundary data

- 13 representative news videos
- Total frames: 597043
- Total transitions: 3785
- Transition types:
  - 1,844 (48.7%) **Cuts** (2005: 60.8%)
  - 1,509 (39.9%) **Dissolves** (2005:30.5%)
  - 51 (1.3%) **Fade-out/-in** (2005: 1.8%)
  - 381 (10.1%) **other** (2005: 6.9%)
- More graduals, which are harder to match

Shot boundary data – more short graduals

- Short graduals: graduals <= 5 frames in length
- Harder to match - treated as “cuts” but no 5-frame expansion as with other cuts to handle differences in decoders
- 2006 data has more “short graduals”

Short graduals	2006	2005	2004	2003
% of graduals	47	35	24	7
% of all	24	14	10	2

Evaluation Measures

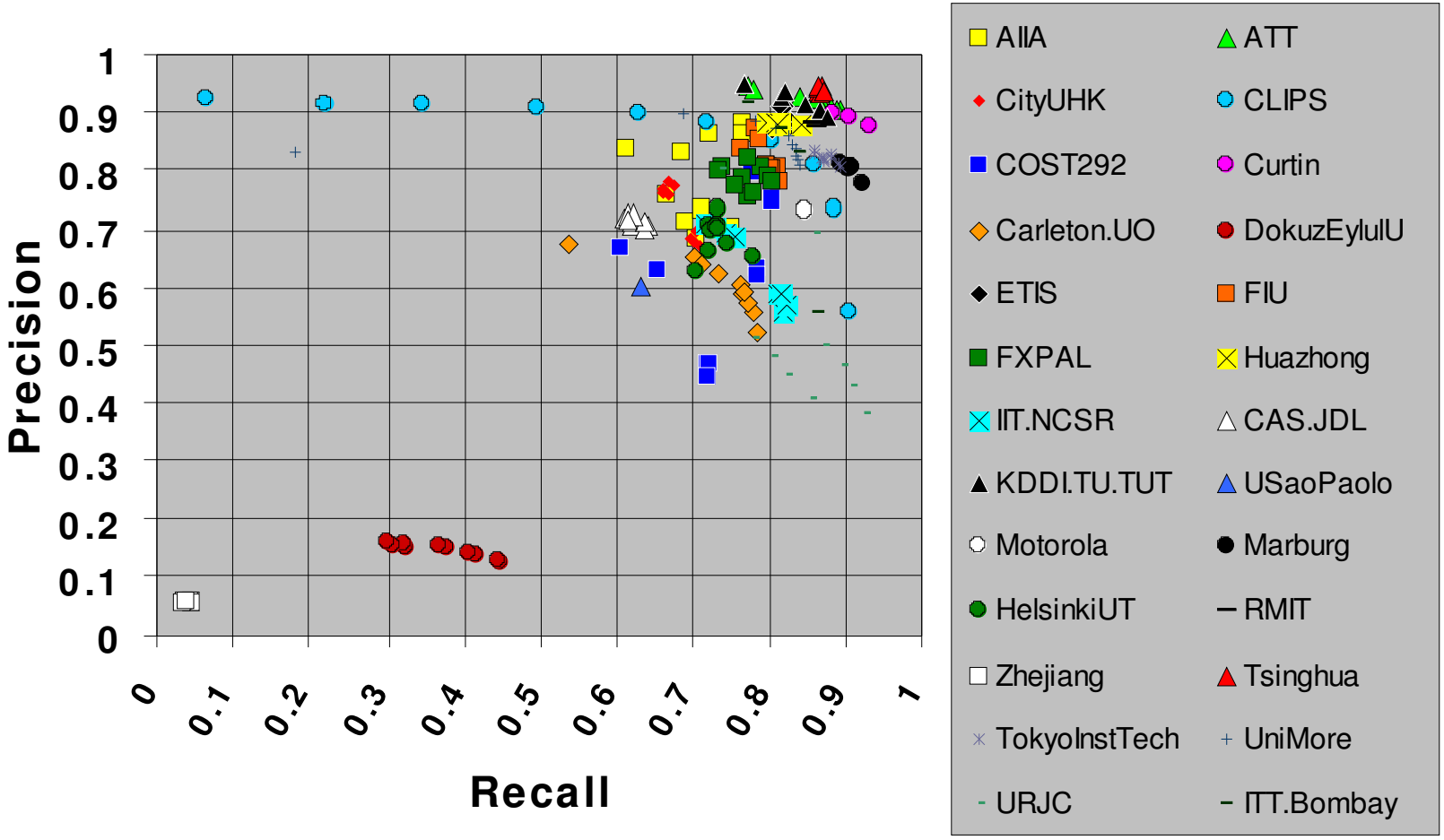
$$\text{Precision} = \frac{\#\ \text{Transitions Correctly Reported}}{\#\ \text{Transitions Reported}}$$

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

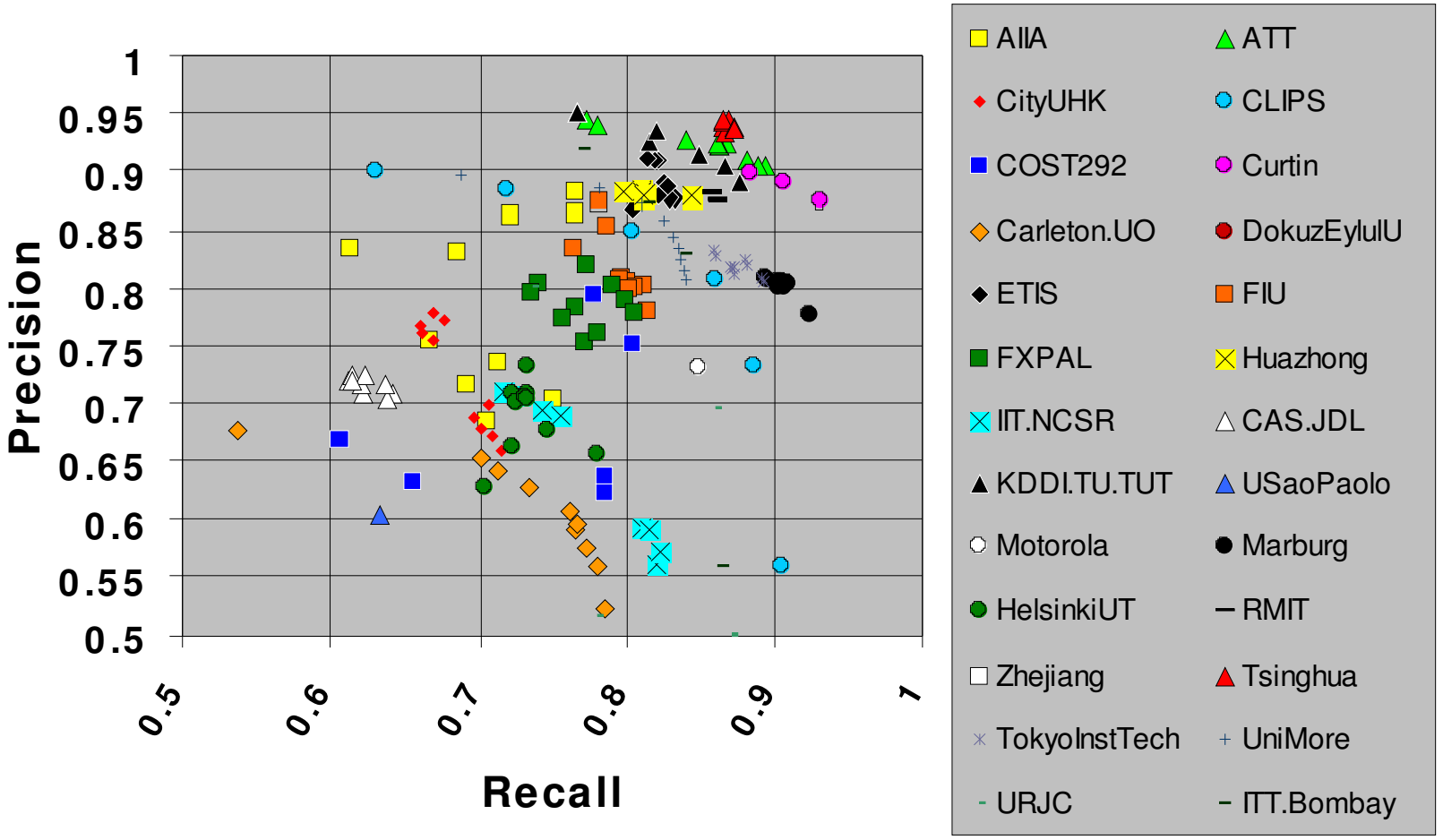
$$\text{Frame Precision} = \frac{\#\ \text{Frames Correctly Reported in Detected Transitions}}{\#\ \text{Frames reported in Detected Transitions}}$$

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

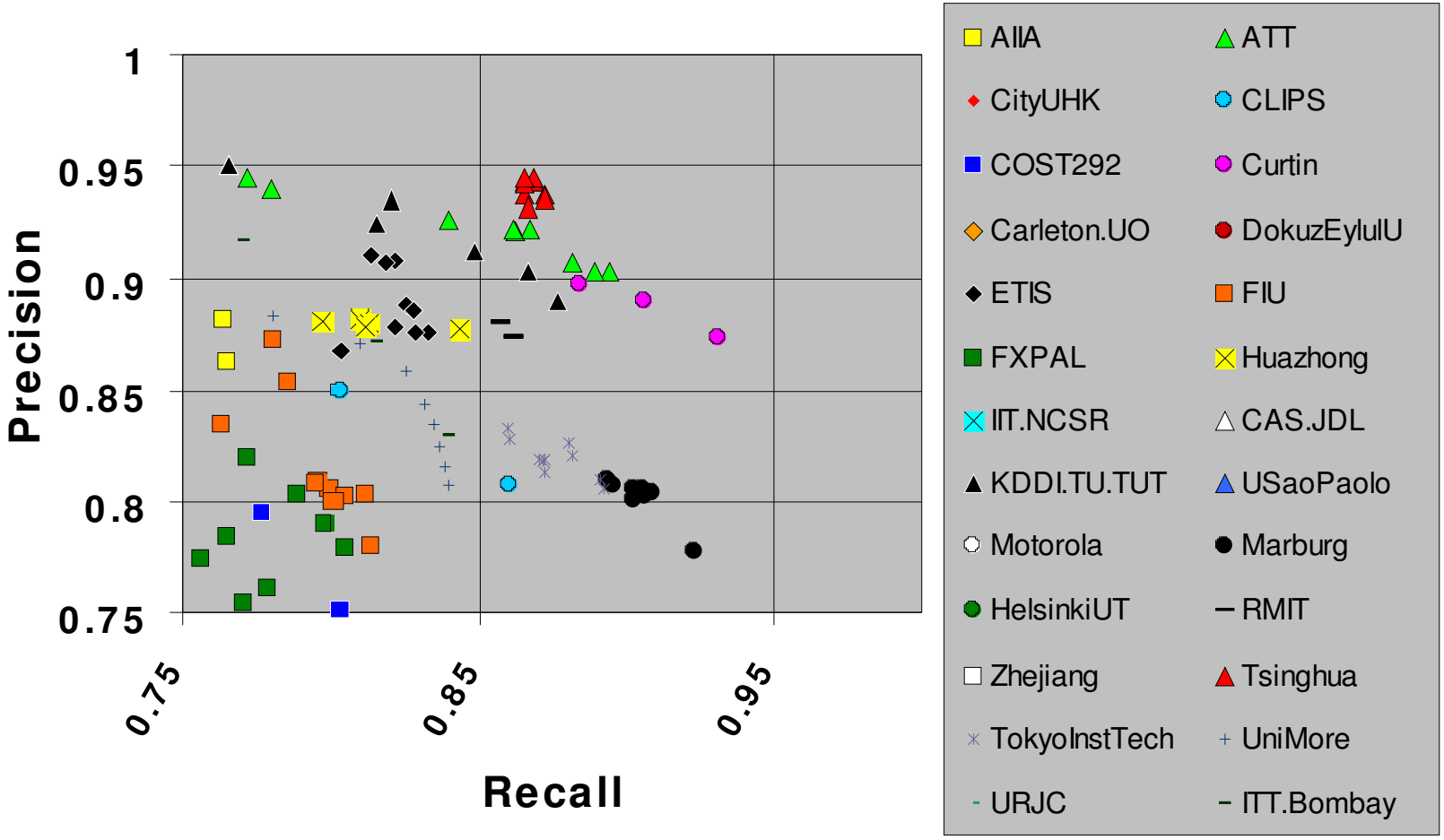
Cuts


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
ATT
CityUHK
CLIPS
COST292
Curtin
Carleton.UO
DokuzEylulU
ETIS
FIU
FXPAL
Huazhong
IIT.NCSR
CAS.JDL
KDDI.TU.TUT
USaoPaolo
Motorola
Marburg
HelsinkiUT
RMIT
Zhejiang
Tsinghua
TokyoInstTech
UniMore
URJC
ITT.Bombay

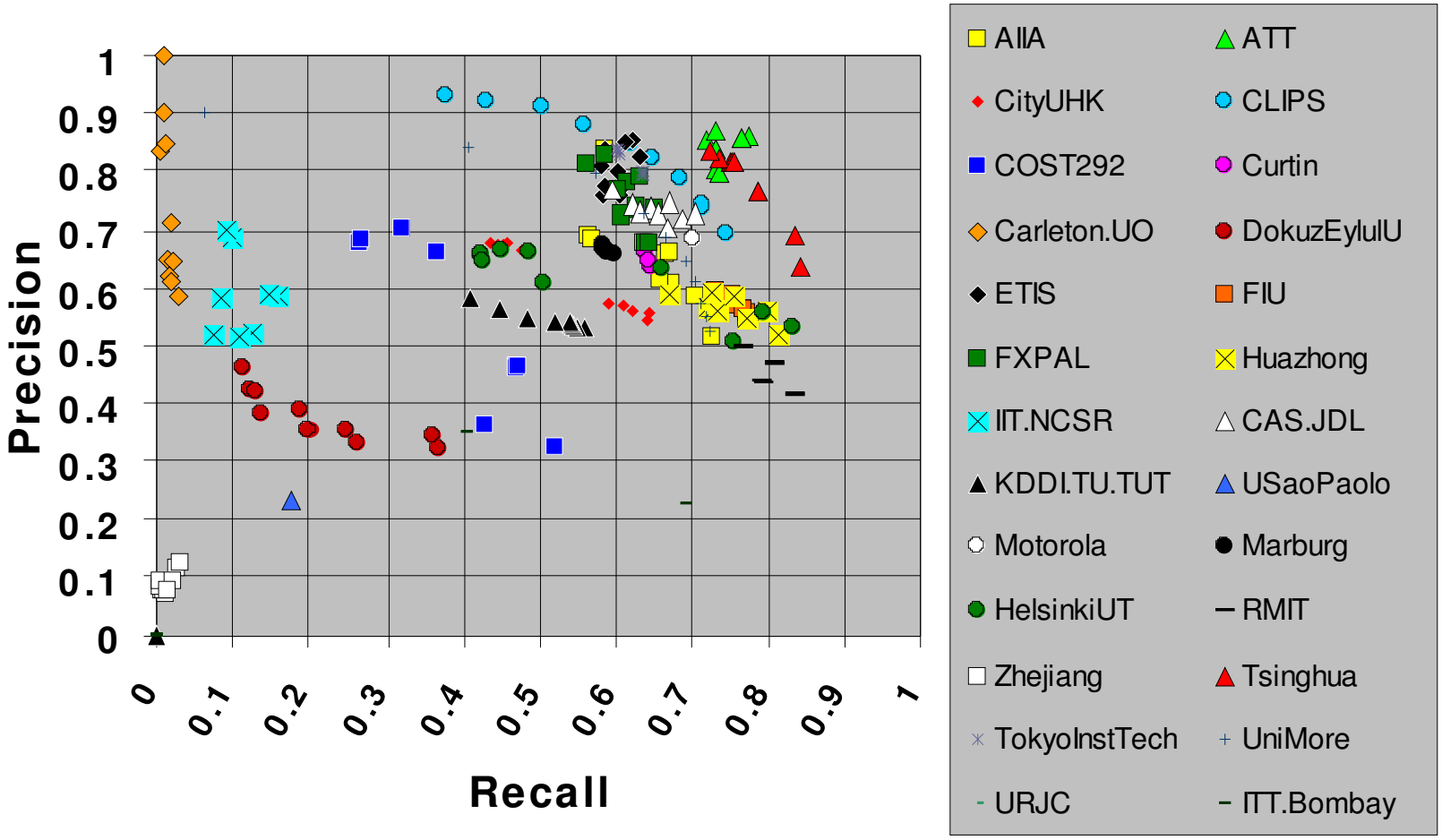
Cuts (zoomed)


Precision
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
AIIA
ATT
CityUHK
CLIPS
COST292
Curtin
Carleton.UO
DokuzEylulU
ETIS
FIU
FXPAL
Huazhong
IIT.NCSR
CAS.JDL
KDDI.TU.TUT
USaoPaolo
Motorola
Marburg
HelsinkiUT
RMIT
Zhejiang
Tsinghua
TokyoInstTech
UniMore
URJC
ITT.Bombay

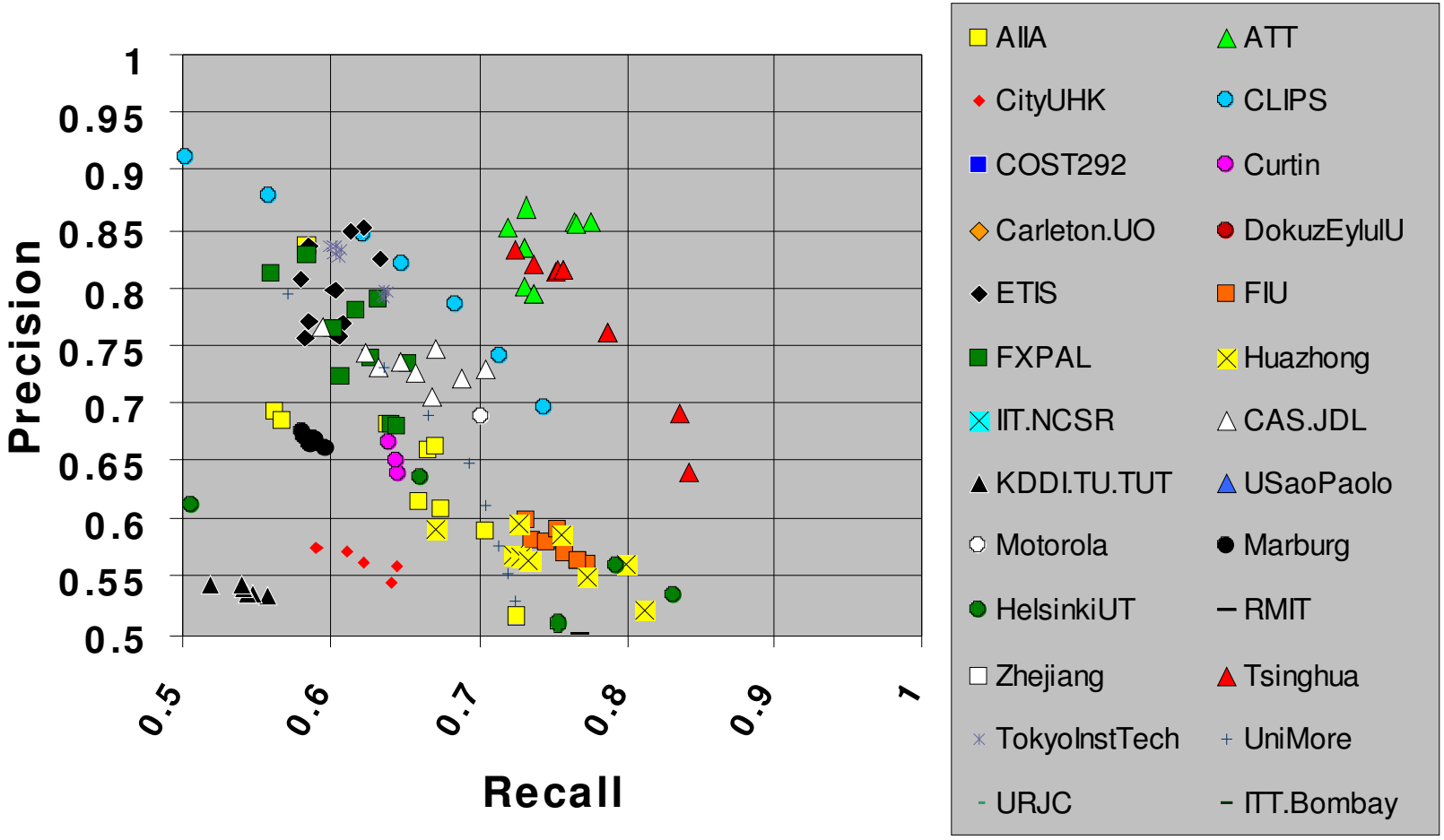
Cuts (zoomed again)



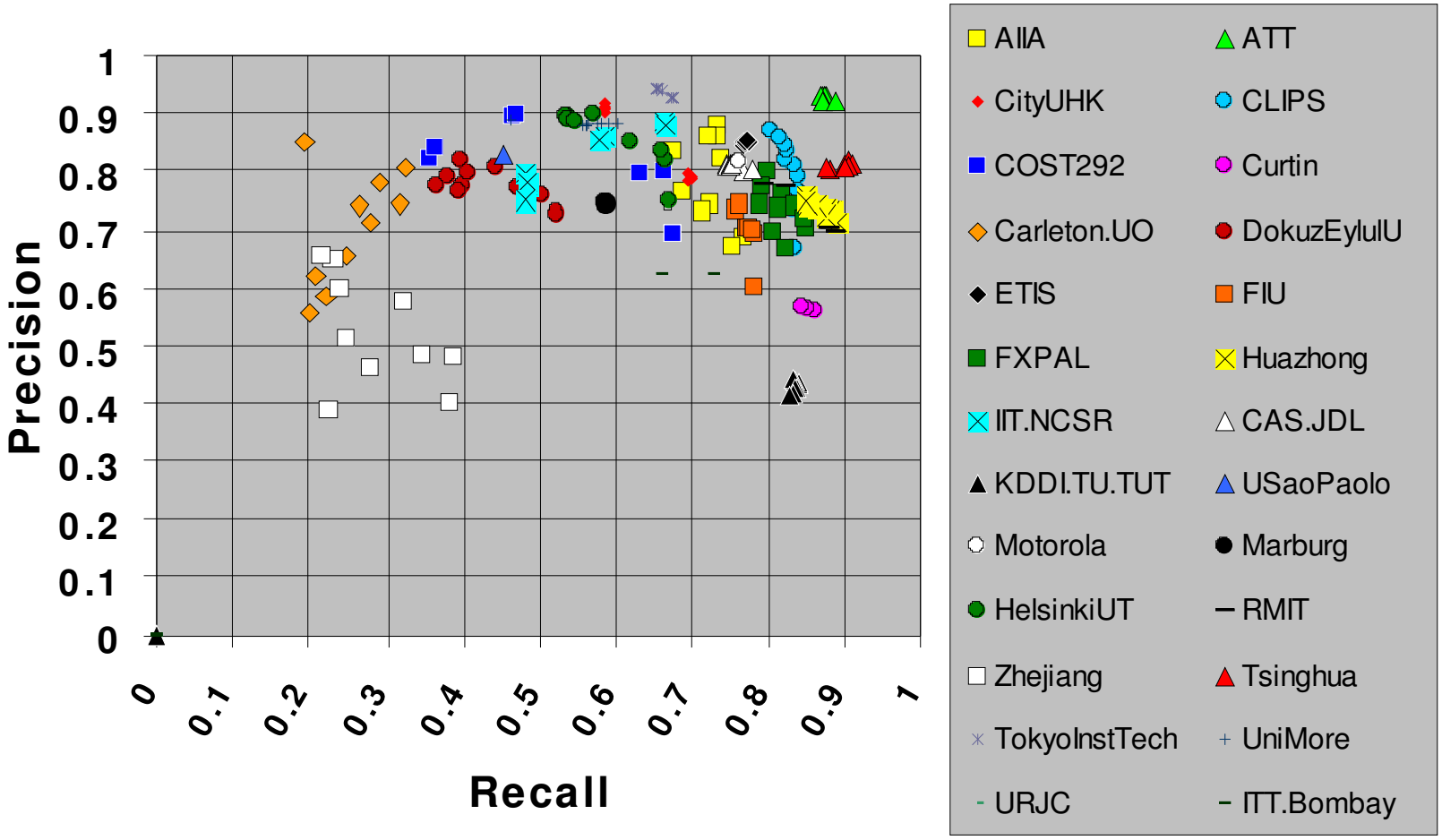
Gradual transitions



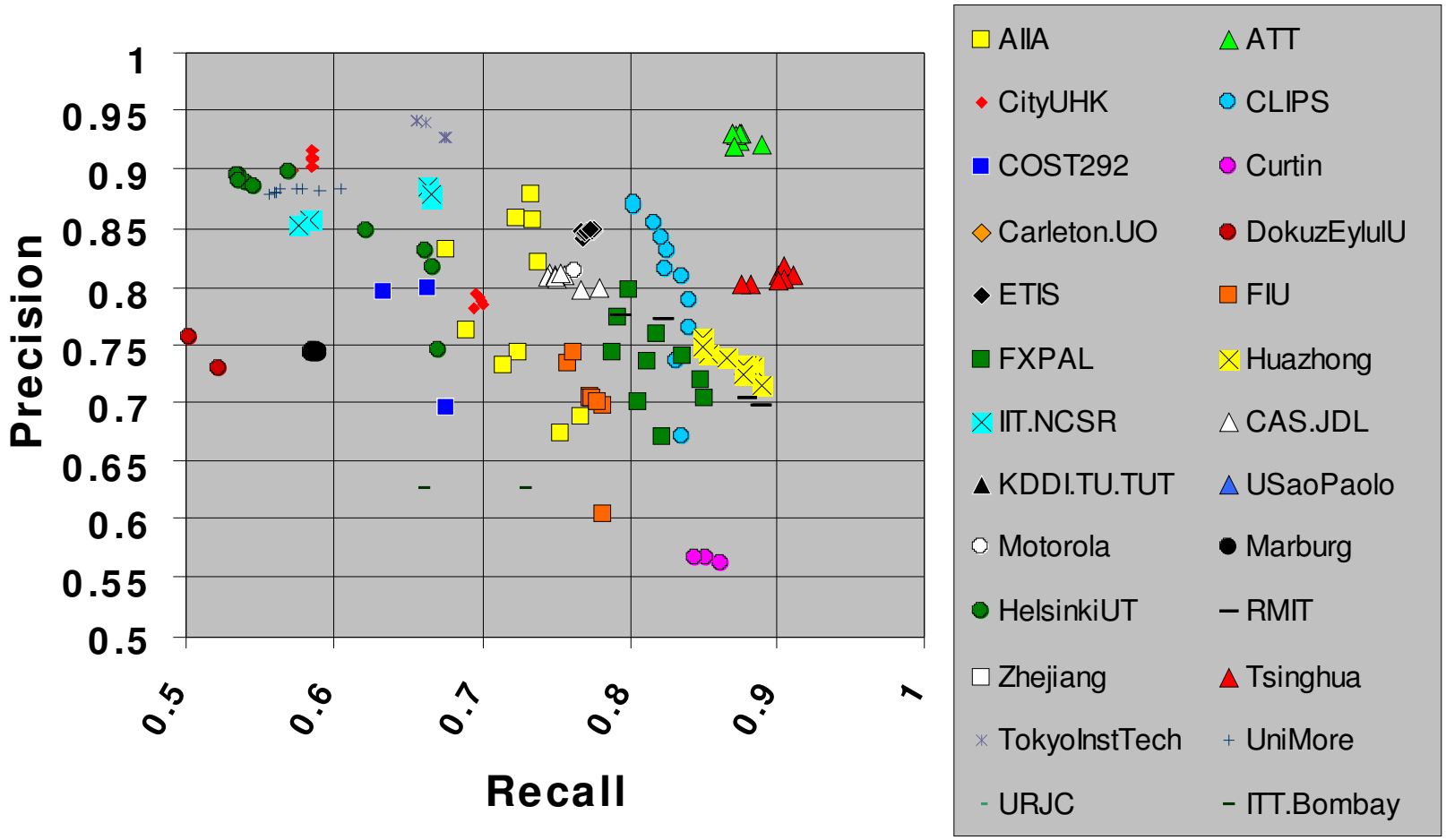
Gradual transitions (zoomed)



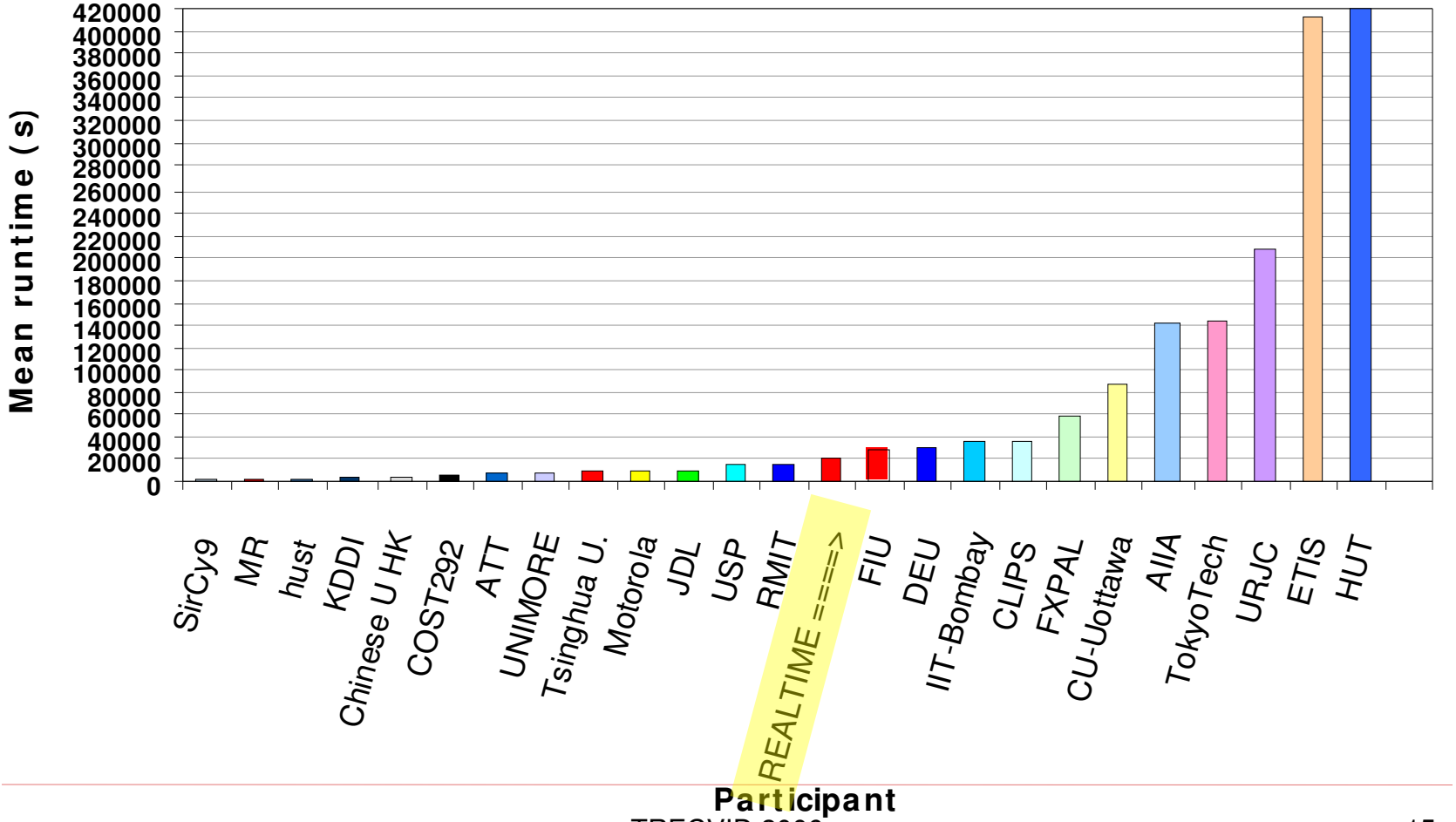
Gradual transitions (Frame-P & -R)



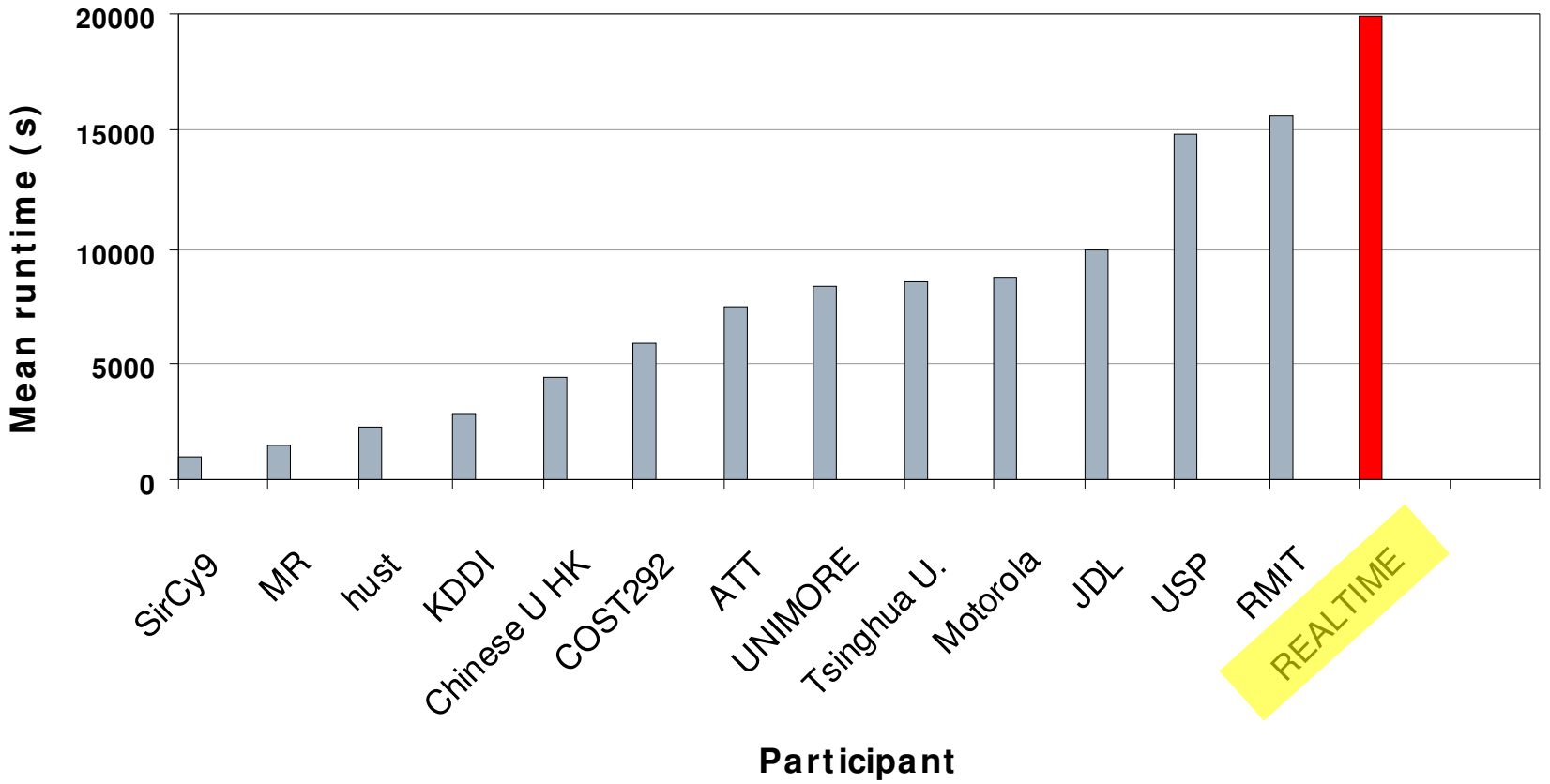
Gradual transitions (Frame-P & -R) zoomed



Mean runtime in seconds

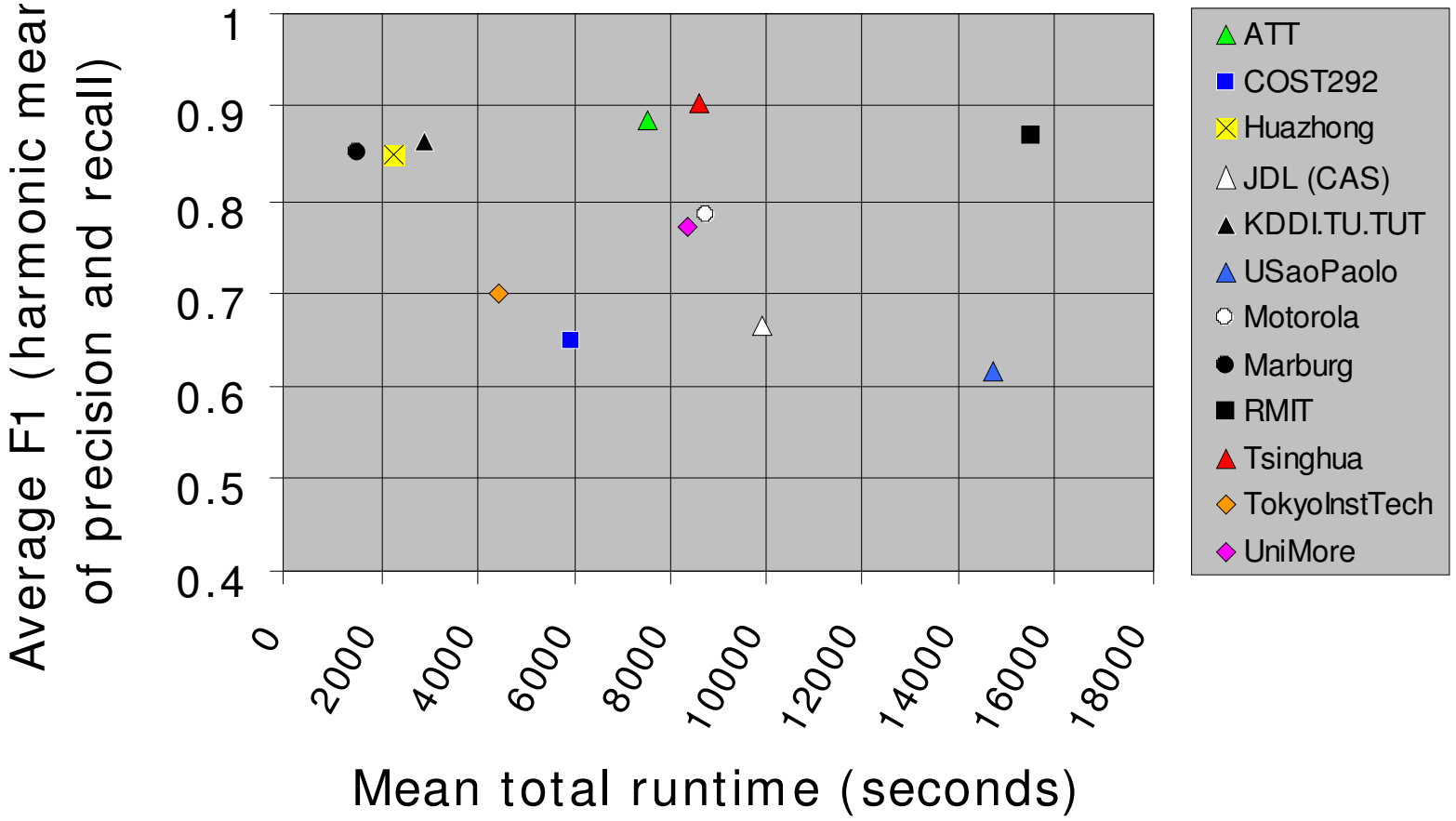


Mean runtime in seconds (faster than realtime)



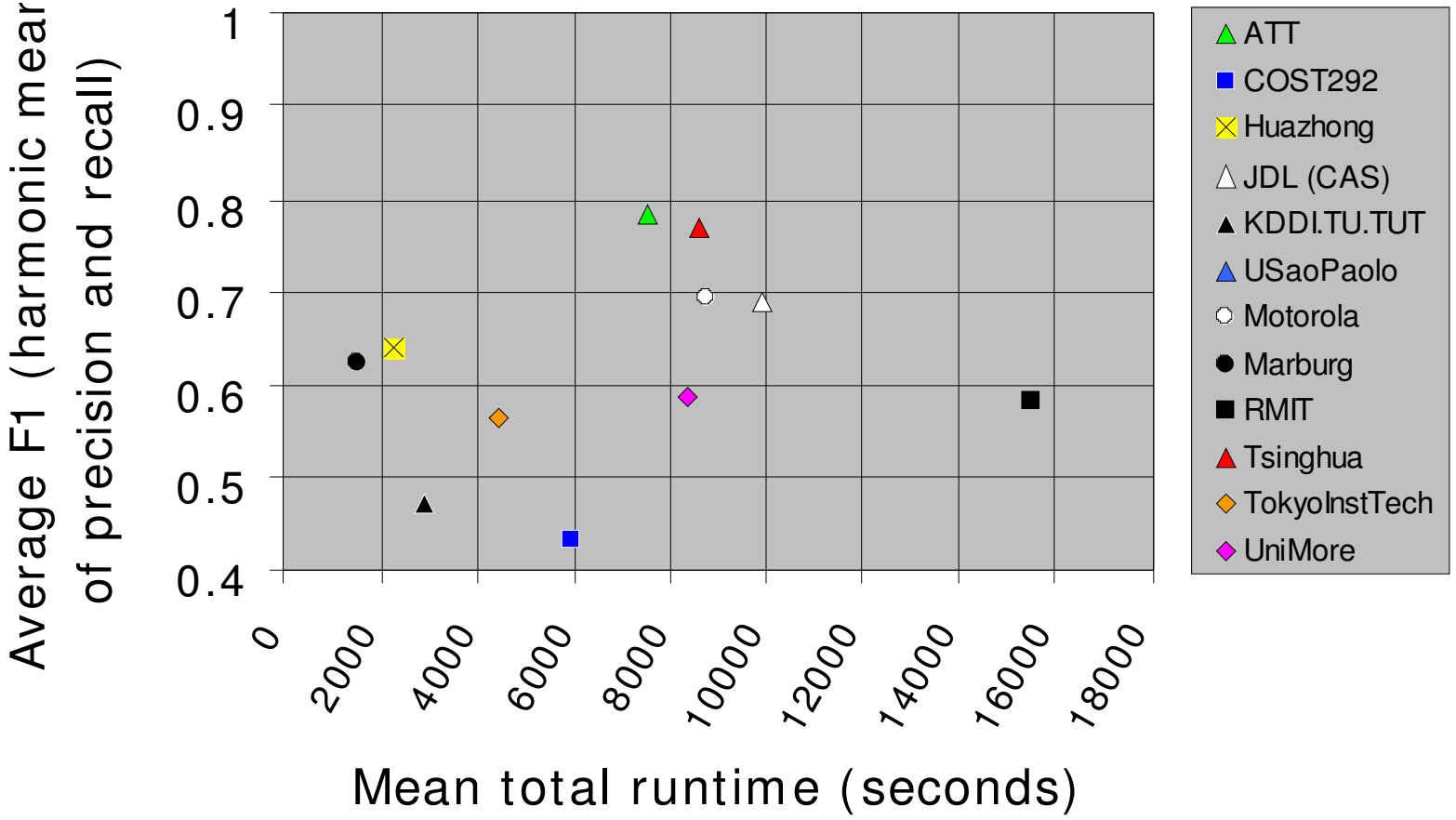
Mean total runtime vs effectiveness on cuts

(for systems faster than realtime)



Mean total runtime vs effectiveness on graduals

(for systems faster than realtime)



1. AIIA Laboratory

- ICASSP2006 paper describes using information from multiple pairs of frames, within a temporal window;
- Good for GTs, which it targets
- 10 runs, varying thresholds
- Frame similarity is color based, not histogram bins but intensity of R,G,B, window size
- Downsampled frame size for 25%
- Performance … several others do better for cuts and also for GTs, but in FR/FP they are better
- Computational expense as expected, several xRT, but novel

2. AT&T Laboratories

- Built 6x independent detectors for cuts, fast dissolves (<5Fs), fade-in, fade-out, dissolve, and wipes;
- Easy to plug in new detectors;
- Fusion of outputs, fuse & resolve conflicts
- Each detector is a FSM (details in paper)
- Extract color RGB & intensity, histograms, edges, average, variance, skew, flatness, all from a central area of frame -> losing the borders;
- Compute frame-frame for adjacent and 6-distant frames;
- Late fusion with prioritisation of detection types;
- 7th fastest in execution and rates well in performance

3. Chinese Academy of Sciences / JDL

- 2-pass approach … histograms and mutual information
- Thresholding to locate possible SBs then a SVM on those candidate areas;
- Rationale based on not needing detailed features around every frame;
- Needs to improve distinction between GTs and camera motion, which gives false +s;
- Histograms are color based
- Results deflated by their decoder being 1 frame out of sync with evaluation numbering;

4. City University of Hong Kong

- Used RGB and HSV color spaces;
- Euclidean distance, color moments and Earth Mover distances
- EMD best
- Used adaptive thresholding, adapting to mean and standard deviations in 11-frame window;
- Good for cuts and short GTs;
- Separate GT detector;

5. CLIPS-IMAG

- Same system as in 2004 and 2005, no new training.
  - *Cut detection by image difference with motion compensation and photographic flash detection.*
  - *GTs by comparing norms of the first and second temporal derivatives of the images.*
- Performance worse than previous years;

6. COST292

- 10 sites, 2 involved in SB task;
- Used existing detectors from TU Delft and from LaBRI U. Bordeaux, merged outputs;
- Delft … spatiotemporal block based analysis based on 3D pixel blocks, not frames or 2D blocks;
- LaBRI is the 2005/2004 detector, improved;
  - *Targets I- and P- frames only, in compressed domain*
- Merging based on intersections and then weighted confidences in each method;
- Submitted both individual and combined runs … combined less than the best individual run;

7. Curtin University

- Late paper ?

8. Dokuz Eylol U.

- Color histograms, Euclidean distance, differences in RGB for frame-frame, with thresholds;
- Used a skip frame interval to skip ahead 5 frames when very similar;
- Big reduction in compute time, small loss in accuracy;
- Effectiveness needs to be improved;

9. Florida International University

- No paper

10. FX Palo Alto Laboratory

- Builds on 2004 and 2005;
- Low level features (global and block colour histograms), feed in to generate mid-level features (interframe similarity matrices), which feeds into a kNN classifier
- Used more favorable training data than previous years “used machine generated output from master shot reference of the development set”

11. Helsinki University of Technology

- Approach is to extract feature vectors from consecutive frames;
- Project these on to a 2D self-organizing map (SOM);
- Detect GTs and cuts from resulting SOM;
- Experimented with cut optimized, GT optimized, blend optimized and different training data sources;
- Computationally the most expensive (because of SOMs);

12. Huazhong U. of Science & Tech.

- No paper

13. Indian Institute of Tecnology, Bombay

- Targets false +ves from dramatic illumination changes (flashes) and shaky camera and fire/explosions;
- Multi-layer filtering to detect candidates based on correlation of intensity features;
- Then use Morlet wavelets to filter candidates and a threshold SVM which uses more detailed features
  - *Pixel differences, color histograms, edges, intensity & wavelets;*
- The best cuts-only and best GTs-only are competitive but the merged combination is not;

14. IIT / NCSR Demokritis

- Spatial Segmentation
- Frame-frame similarities between consecutive frames using Earth Mover's distance;
- Combination of RGB color, adjacent RGB color, center of mass and adjacent gradients;
- Independent modeling and detection of cuts and GTs;
- Hard cuts OK, GTs weak -- plan to include motion information;

15. KDDI / Tokushima U. / ISM / NII

- Very fast execution time and among best performances;
- Extension of 2005 approach and new detection of long dissolves;
- 2-stage SVMs with combination of multi-kernals
- Features used are:
  - *Number of in-edges, number of out-edges;*
  - *Pixel intensities;*
  - *FX-PAL 2004 approach;*
  - *Edge change ratio;*

16. ETIS

- SVMs as standard trained classifiers;
- Independent cut and GT detectors;
- CUTS - features are color histograms, variations on moments for shape description, projection histograms,
- GTs - features are illumination variations and global edge information;
- Also includes a fade detector;
- Trained on Brazilian TV commercials, only 2 min and 2 sec of it ?
- Computationally the most expensive;

17. Motorola Multimedia Research Lab.

- No paper

18. RMIT University

- Building on previous TRECVids
- Based on a moving query window yet performance is approx real time;
- Performance in 2006 is less than previous years, possibly because of harder data, especially on GTS.
- HSV color bins for regions of the frame, with weightings for some regions;

19. Tokyo Institute of Technology

- No paper

20. Tsinghua University

- Same system as TRECVid 2005 but improved;
- Ran 2006 system on 2005 data yielding better performance than 2005, so system better;
- Yet 2006 figures are worse than 2005 figures --> data is officially harder
- Improvements are in the detection of FOIs, flashes and short GTs;
- Uses an FOI detector, independent CUT and GT detection, and targets the transitions in video-in-video, which are not SBs;
- Possibly the best performance and again, very fast;

21. University of Marburg

- Unsupervised k-means clustering for Cuts and GTs, extending TRECVid2005 system;
- Cuts …
  - *2 different frame dissimilarity measures namely motion-compensated pixel differences and color based histograms*
- GTs …
  - *Dissimilarities for different frame distances, same dissimilarity measures as cuts. Explicit fade detector;*
- Good for cuts … execution performance ?
- Unsupervised approach … “reached a level of robustness and detection quality … (especially) for cuts”

22. University of Modena Reggio

- Follows TRECVid in 2005 (with FSU)
- Targets GTs which have linear frame transitions, but it also works for cuts;
- Work on determining the range (in frames) and nature of a GT and integrating Cut and GT detectors;
- Works on windows of 60 frames;
- Not clear what (frame) similarity is used;
- Quite fast;

23. Carleton University (Ottawa)

- Approach based on tracking image features across frames, and if a lot of features drop off in the tracking, then likely shot bound;
- Designed for non-news video … movies, TV, etc.
- “features” are corners of edges on the greyscale frames;
- Requires registration of corner features across frames;
- Needs automatic thresholding to adjust to video type;
- Inherently computationally very expensive, but includes some “tricks” to reduce time, but still 5x RT at least;
- Very different;

24. University of Sao Paulo (USP)

- 2-step process
  - *Compute absolute pixel differences between adjacent frames to detect 'events' … any type of large discontinuity or activity in pixels;*
  - *Histogram intersection difference on candidate areas from (1);*
- Designed for cuts only;

25. University Rey Juan Carlos

- Builds on TRECVid 2005, fusing color and shape primitives;
- Color == 16-bin histogram;
- Shape == Zernike moments;
- Varied the weighed combinations and found a fusion approach that improved on the independents in isolation;
- Computation of Zernike moments can be expensive;

- Interesting results of 2006 system on 2005 and 2006 data showed 2006 data much poorer performance;

26. Zhejiang University

- Fastest performance but some programming error in cuts, GTs are better

- Paper doesn’t say they did SBD !

Observations

- Excellent performance on cuts and graduals **despite more difficult data**
- Good effectiveness achievable at significantly less than realtime
- Despite the continued introduction of novel approaches, novelty =/= improvement
- Interest in the task seems strong … but ..

- Seems time to retire this task, what more can we learn ?