# TRECVID-2008 Content-based Copy Detection task Overview

Wessel Kraaij
TNO // Radboud University

George Awad, Paul Over
NIST

# Outline

- Task overview
- Dataset and queries
- Transformations
- Evaluation metrics
- Participants
- Results
- Global Observations
- Issues

# Task design considerations

- Copy detection is applied in several real-word tasks:
  - television advertisement monitoring
  - detection of copyright infringement
  - detection of known (illegal) content
- 2008: pilot task with synthetic queries
- Audio handled in a separate condition
- Task has both a detection and localization component
- Detection measure based on error rates
- Weighted trade-off of type I (false alarms) and type II (misses) errors.
- Computation of optimal operating point by NIST, next time by participants?

# CBCD task overview

- Goal:
  - Build a benchmark collection for video copy detection methods
- Task:
  - Given a set of reference (test) video collection and a set of 2000 queries,
  - determine for each query if it contains a copy, with possible transformations, of video from the reference collection,
  - and if so, from where in the reference collection the copy comes
- Three main task types were derived:
  - Copy detection of video-only queries (required)
  - Copy detection of audio-only queries (optional)
  - Copy detection of video + audio queries (optional)

# Datasets and queries

- Dataset:
  - Reference video collection: TV2007 and TV2008 sound & vision data (~200 hr)
  - Non-reference video collection : TV2007 BBC rushes data
- Query types: (Developed by INRIA-IMEDIA)

Copies {
  - Type 1: composed of a reference video only. (1/3)
  - Type 2: composed of a reference video embedded in a non-reference video. (1/3)
  - Type 3: composed of a non-reference video only. (1/3)

- Number of queries:
  - 201 total original queries were created by NIST using tools created by INRIA-IMEDIA
  - 67 queries for each type
- After creating the queries, each was transformed.
  - 10 video transformations by Laurent Joyeaux (independent agent at INRIA)
  - 7 audio transformations by Dan Ellis (Columbia University)
- Yielding…
  - 10 * 201 = 2010 video queries
  - 7 * 201 = 1407 audio queries
  - 10 * 7 * 201 = 14070 audio+video queries

# Video transformations

- Cam Cording (T1)
- Picture in picture (T2)
- Insertions of pattern (T3)
- Strong re-encoding (T4)
- Change of gamma (T5)
- Decrease in quality (T6, T7) – by introducing a combination of *Blur, Gamma, Frame dropping, Contrast, Compression, Ratio, White noise*
  - For T6, 3 transformations are randomly selected and combined
  - For T7, 5 transformations are randomly selected and combined
- Post production (T8, T9) – by introducing a combination of *Crop, Shift, Contrast, Text insertion, Vertical mirroring, Insertion of pattern, Picture in picture,*
  - For T8, 3 transformations are randomly selected and combined
  - For T9, 5 transformations are randomly selected and combined
- Combination of 5 randomly selected transformations chosen from T1-9 (T10)

# Video transformations examples

Picture in Picture

Blur


Insertion of pattern

Strong re-encoding

Noise

Contrast

Change in gamma

Mirroring

Ratio

Crop

Shift


Text insertion

# Some actual query clips






# Audio transformations

- T1: nothing
- T2: mp3 compression
- T3: mp3 compression and multiband companding
- T4: bandwidth limit and single-band companding
- T5: mix with speech
- T6: mix with speech, then multiband compress
- T7: bandpass filter, mix with speech, compress

# Some important task details/assumptions

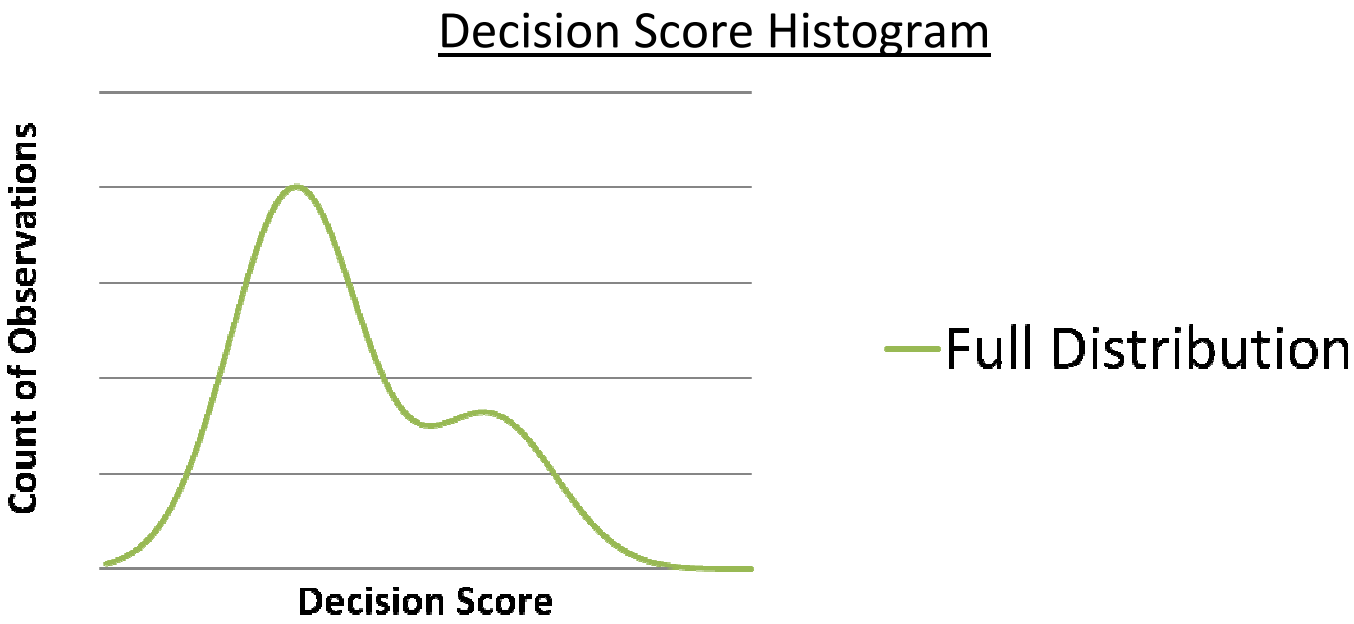
- Detection systems use a form of thresholding
- Systems are asked to output a list of possible copies (in fact disabling thresholding)
- A query can yield just one true positive
- A query can give rise to many false alarms
- Consequence:
  - Type I error modeled as *false alarm rate*
  - Type II error modeled as *Pmiss*

# Evaluation metrics

- Three main metrics were adopted:
  1. Normalized Detection Cost Rate (NDCR)
     - measures error rates/probabilities on the test set:
       - Pmiss (probability of a missed copy)
       - Rfa (false alarm rate)
     - combines them using assumptions about one possible realistic scenario:
       - *Copy target rate (Rtarget)* = 0.5/hr
       - *Cost of a miss (CMiss)* = 10
       - *Cost of a false alarm (CFA)* = 1
  2. $F_1$ (how accurately the copy is located, harmonic mean of P and R)
  3. Mean processing time per query
- General rules:
  - No two query result items for a given video can overlap.
  - For multiple result items per query, one mapping of submitted extents to ref extents is determined based on F1-score.
- The reference data has been found if and only if:
  - The asserted test video ID is correct AND asserted copy and ref. video overlap.

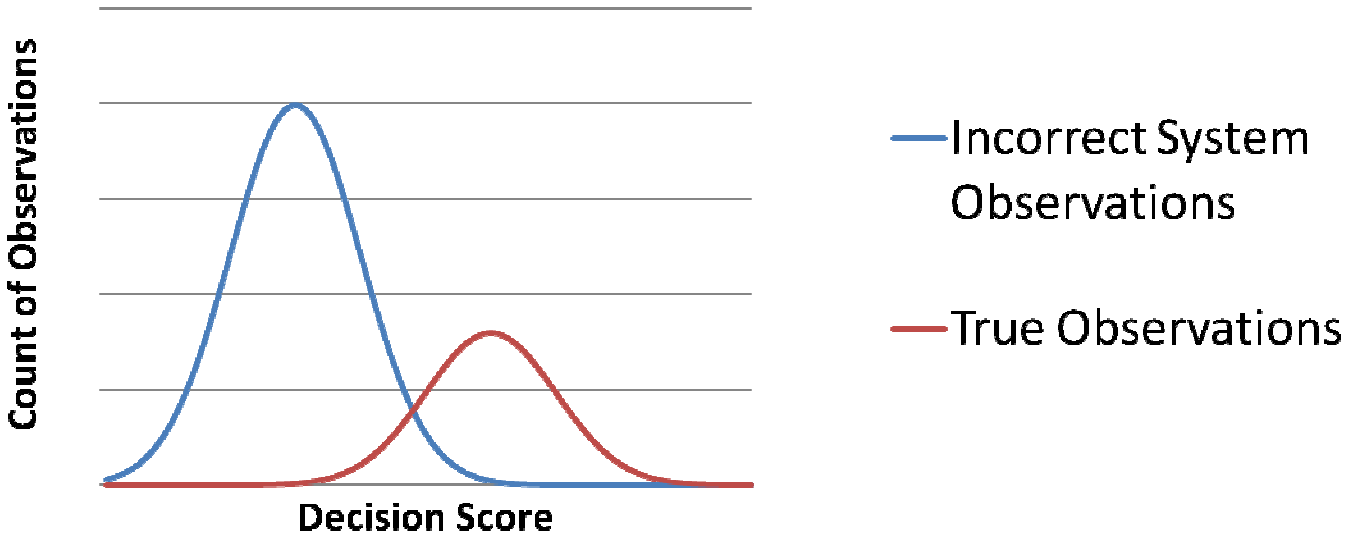
[Kraaij, Over, Fiscus, Joly,2008] Final CBCD Evaluation plan TRECVID 2008 v1.3

# Decision Error Tradeoff Curves $Prob_{Miss}$ vs. $Rate_{FA}$



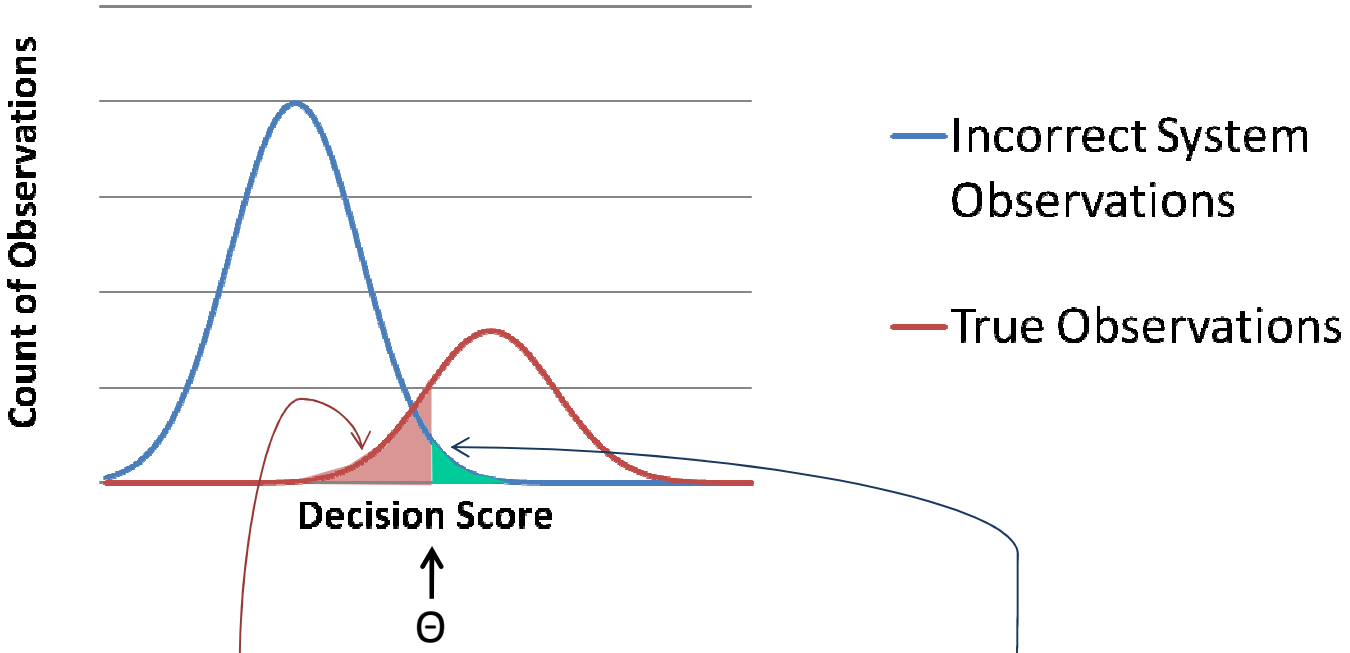
# Decision Error Tradeoff Curves $Prob_{Miss}$ vs. $Rate_{FA}$

Decision Score Histogram Separated wrt. Reference Annotation s



# Decision Error Tradeoff Curves $Prob_{Miss}$ vs. $Rate_{FA}$

Decision Score Histogram Separated wrt. Reference Annotation s

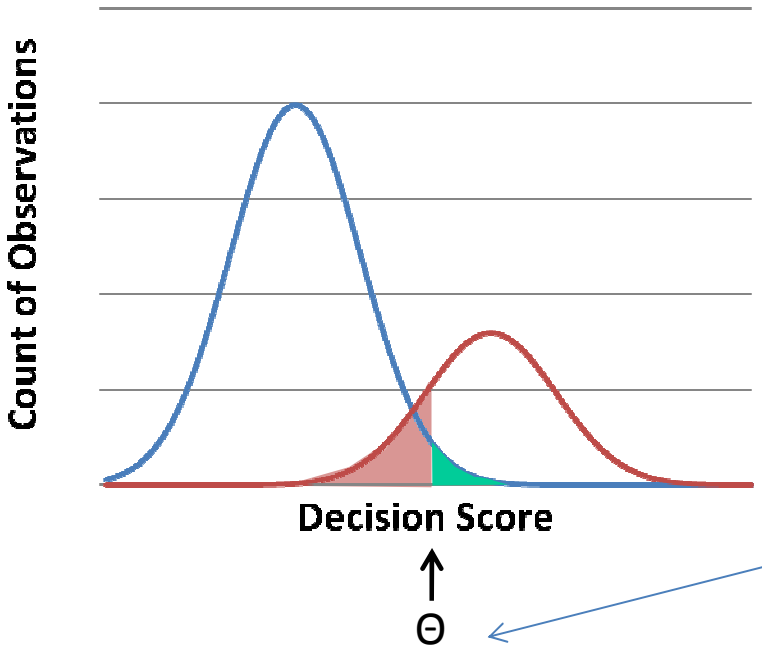


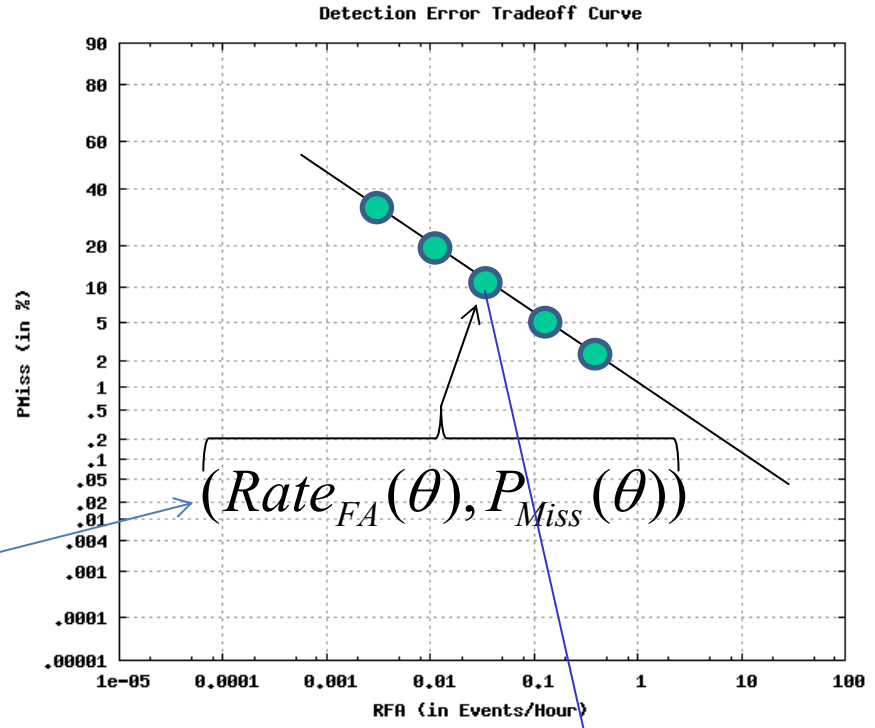

$$P_{Miss}(\theta) = \frac{\#MissedObs}{\#TrueObs} \qquad Rate_{FA}(\theta) = \frac{\#FalseAlarms}{SignalDuration}$$

signal: query

# Decision Error Tradeoff Curves $Prob_{Miss}$ vs. $Rate_{FA}$

Compute $Rate_{FA}$ and $P_{Miss}$ for all Θ




$NDCR = Pmiss + \beta\ Rfa$

$\beta$ *defined by task characteristics*

Optimal threshold determined by NIST

# 22 Participants (finishers)

| | |
|---|---|
| Bilkent University | CD -- FE -- ** |
| University of Bradford | CD ** -- RU -- |
| Beijing Jiaotong University | CD ** ** -- ** |
| Brno University of Technology | CD ED FE ** SE |
| Beijing University of Posts and Telecommunications | CD ** FE -- -- |
| Columbia University | CD -- FE -- SE |
| Computer Research Institute of Montreal | CD -- -- -- -- |
| COST292 Team (Delft Univ.) | CD ** FE RU SE |
| Fudan University | CD ED FE -- SE |
| IBM T. J. Watson Research Center | CD ** FE ** SE |
| INRIA-LEAR | CD -- FE -- -- |
| INRIA-IMEDIA | CD -- ** -- ** |
| Istanbul Technical University | CD -- -- -- -- |
| Chinese Academy of Sciences (MCG-ICT-CAS) | CD ED FE -- SE |
| National Institute of Informatics | CD ** FE RU SE |
| Orange Labs - France Telecom Group | CD -- -- -- -- |
| PicSom(Helsinki University of Technology) | CD -- FE RU SE |
| Tsinghua University – Intel China Research Center | CD ** FE RU SE |
| TNO-ICT | CD ** -- -- -- |
| University of Glasgow | CD -- ** RU SE |
| VIREO (City University of Hong Kong) | CD ** FE RU SE |
| vision@ucf (University of Central Florida) | CD ED ** -- ** |

** : group didn't submit any runs

-- : group didn't participate

# Submission types

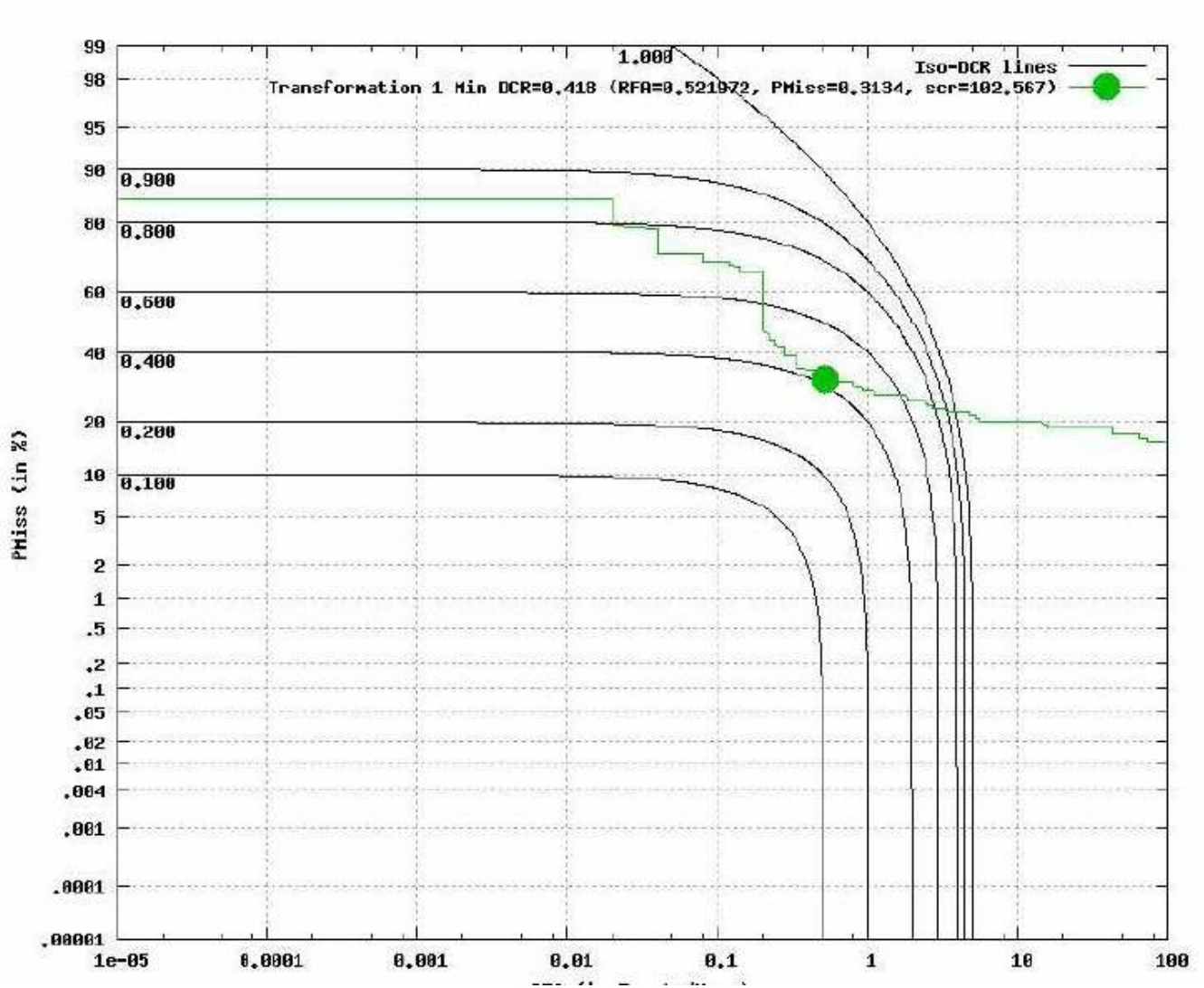
| Run type | Count |
|---|---|
| V (video only) | 48 |
| A (audio only) | 1 |
| M (video + audio) | 6* |
| Total runs | 55 |

*) 1 of these is audio only

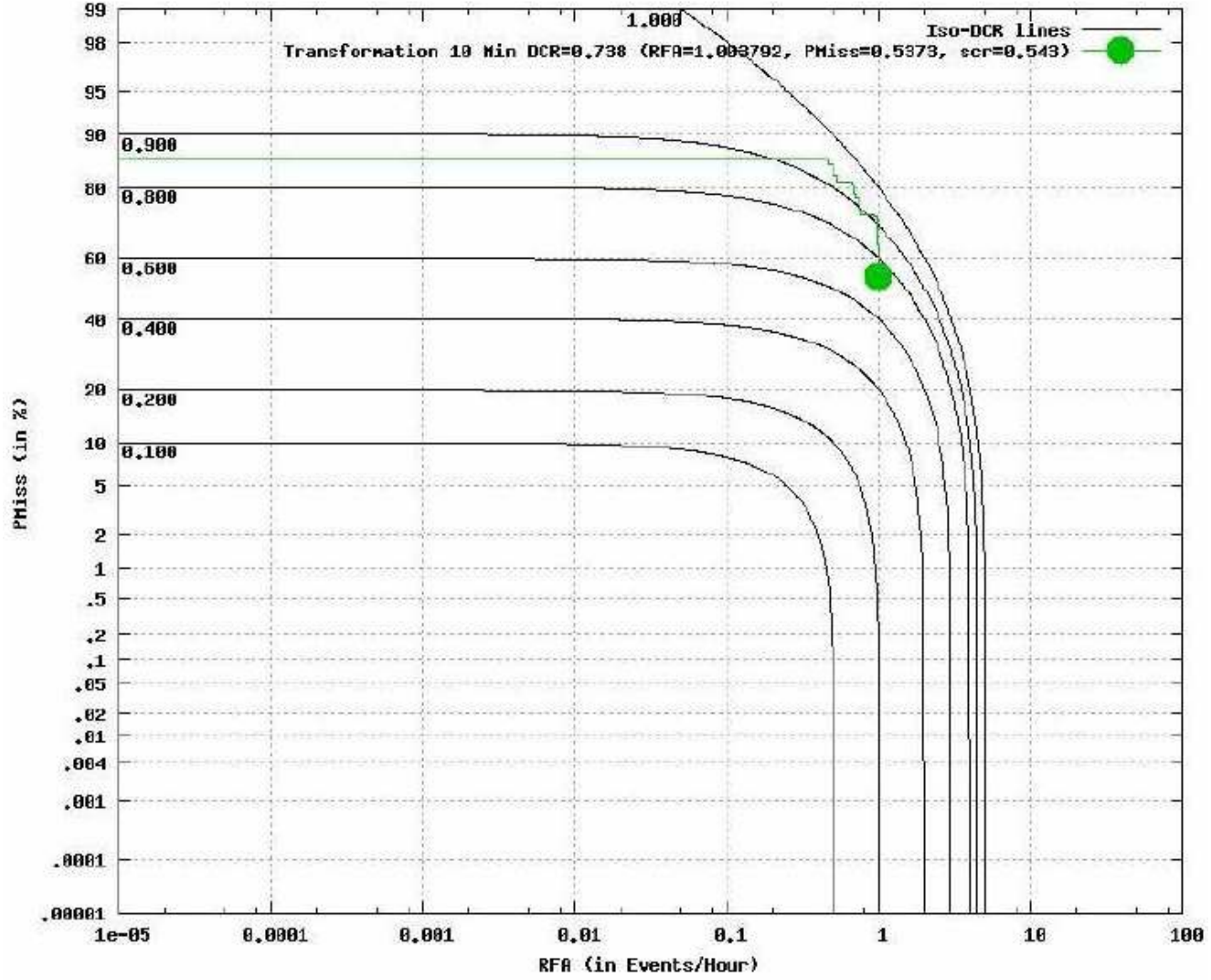
# Approaches

- Typical outline
  - Keyframe extraction
  - Feature extraction
  - Indexing
  - Matching
- Frame representations
  - SIFT descriptors
  - Block based features
  - Global (edge histogram)
  - Differ in efficiency / effectiveness trade-off
- Several teams created a validation/development set
- Different modules for different transformation types vs. generic approaches
- Application of video similarity component developed for rushes summarization
- Combination of audio and video: apply AND, confidence score normalization (training data necessary).

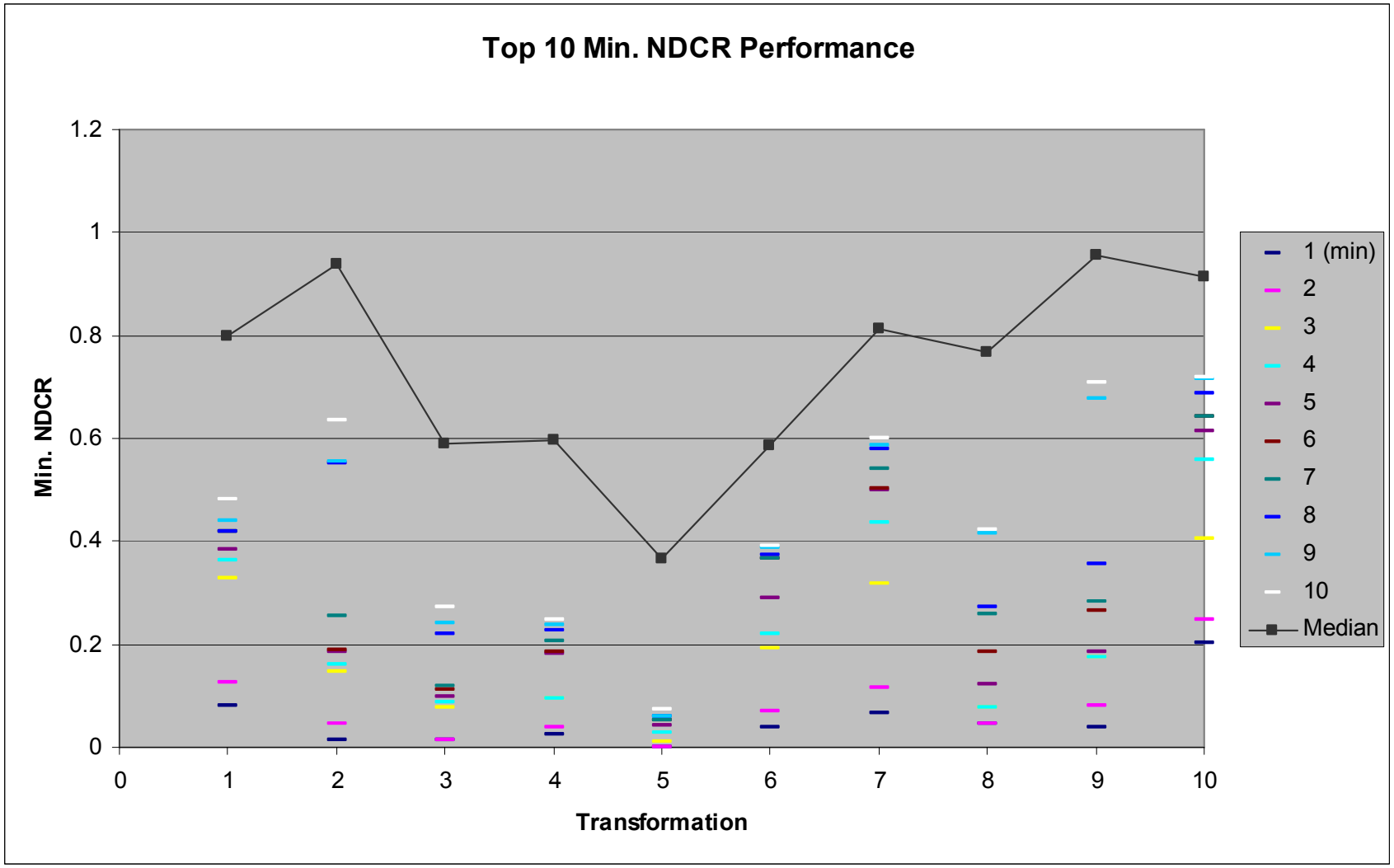
# Example det curve1: Determining the optimal NDCR



# Example det curve2: Cut-off too early?



# CBCD evaluation (Top 10 performance)




T1: Cam Cording *T3: Insertion of patterns* *T5: Change of gamma* T8, T9: Post Production

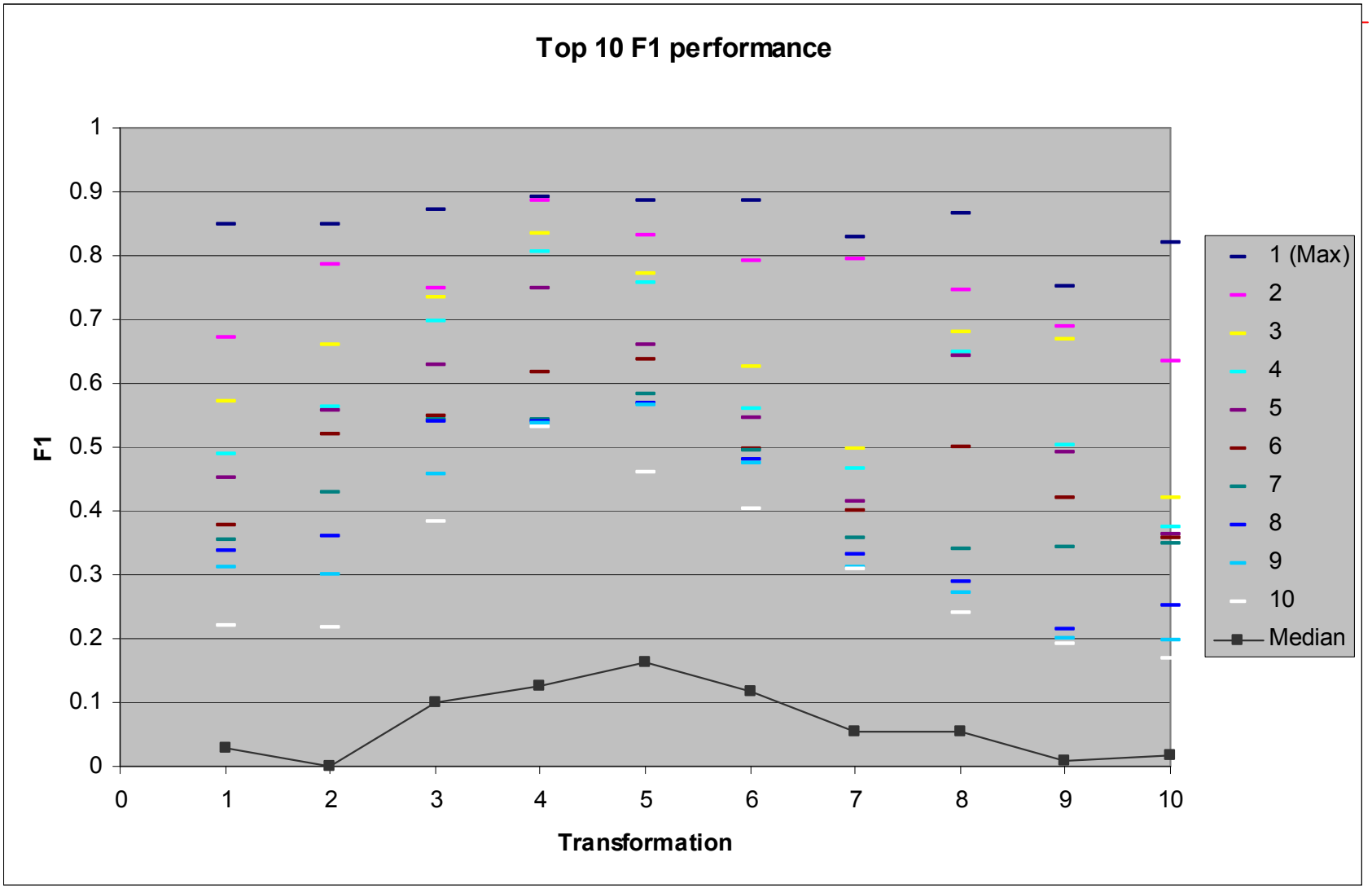
T2: Pict. In Pict. *T4: Re-encoding* T6, T7: Decrease in quality T10: Random combination of 5 transformations

# Top 10 sites per transformation (Min. NDCR)

| Site Name | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T10 |
|---|---|---|---|---|---|---|---|---|---|---|
| INRIA-LEAR.v.Strict | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 |
| INRIA-LEAR.v.Soft | 2 | 2 | 1 | 2 | 3 | 2 | 2 | 1 | 2 | 2 |
| INRIA-IMEDIA.v.fusion | 4 | 4 | 3 | 3 | 1 | 3 | 6 | 6 | 7 | 4 |
| INRIA-LEAR.v.KeysAdves | 3 | 7 | 8 | 7 | - | 5 | 3 | 7 | 6 | 3 |
| INRIA-IMEDIA.v.joly | 9 | - | 4 | 4 | 7 | 4 | 5 | - | - | 5 |
| OrangeLabsICM.v.Run2 | 10 | 6 | 6 | - | 9 | - | - | 4 | 5 | 7 |
| OrangeLabsICM.v.Run1 | - | 5 | 5 | - | 4 | - | - | 3 | 3 | 6 |
| OrangeLabsICM.v.Run3 | - | 3 | 7 | - | 5 | - | - | 5 | 4 | 8 |
| MCG-ICT-CAS.v.ICTCBCDREL | - | 9 | - | 9 | - | 7 | 7 | 9 | - | 10 |
| INRIA-IMEDIA.v.ViCopT | 6 | - | 9 | - | 10 | - | - | 8 | 8 | - |
| MCG-ICT-CAS.v.ICTCBCDALL | - | 8 | - | - | - | - | 10 | 10 | 10 | - |
| CRIMontreal.v.Run1 | 5 | - | - | - | - | 8 | 4 | - | - | - |
| thu-intel.v.2 | - | - | - | 5 | 8 | 9 | - | - | - | - |
| thu-intel.v.3 | - | - | - | 6 | 6 | 10 | - | - | - | - |
| BeijingUPT.v.run1 | - | - | 10 | 8 | - | 6 | - | - | - | - |
| CRIMontreal.v.Run2 | 7 | - | - | - | - | - | 9 | - | 9 | - |
| CRIMontreal.v.Run2Faster | 8 | - | - | - | - | - | 8 | - | - | - |
| ColumbiaU.v.baseLocal | - | - | - | - | - | - | - | - | - | 9 |
| thu-intel.v.1 | - | - | - | 10 | - | - | - | - | - | - |
| MCG-ICT-CAS.v.ICTCBCDTOA | - | 10 | - | - | - | - | - | - | - | - |

*Numbers in table represent the rank

# CBCD evaluation (Top 10 performance)




T1: Cam Cording

T2: Pict. In Pict.

T3: Insertion of patterns

T4: Re-encoding

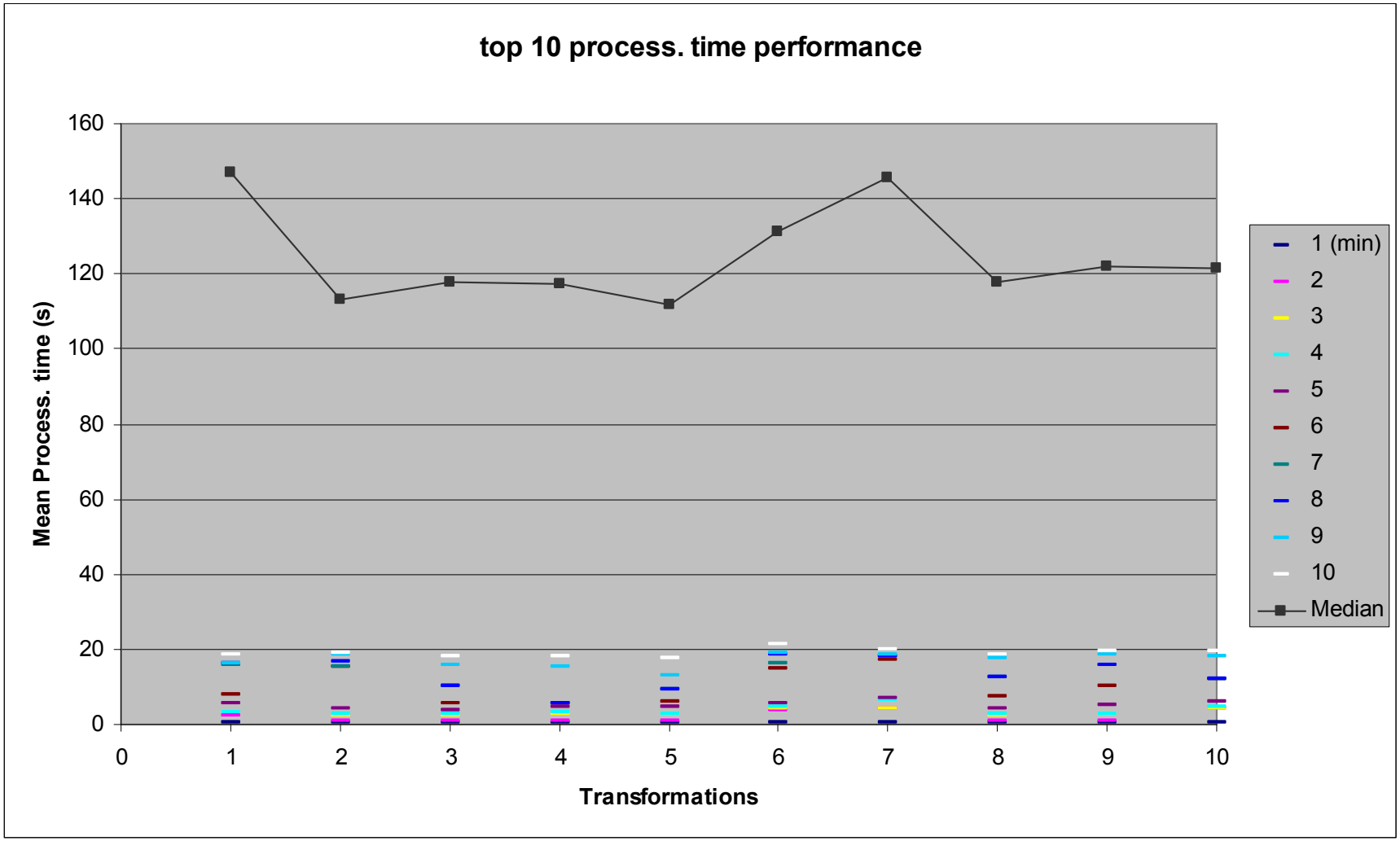
T5: Change of gamma

T6, T7: Decrease in quality

T8, T9: Post Production

T10: Random combination of 5 transformations

# CBCD evaluation (Top 10 performance)




T1: Cam Cording

T2: Pict. In Pict.

T3: Insertion of patterns

T4: Re-encoding

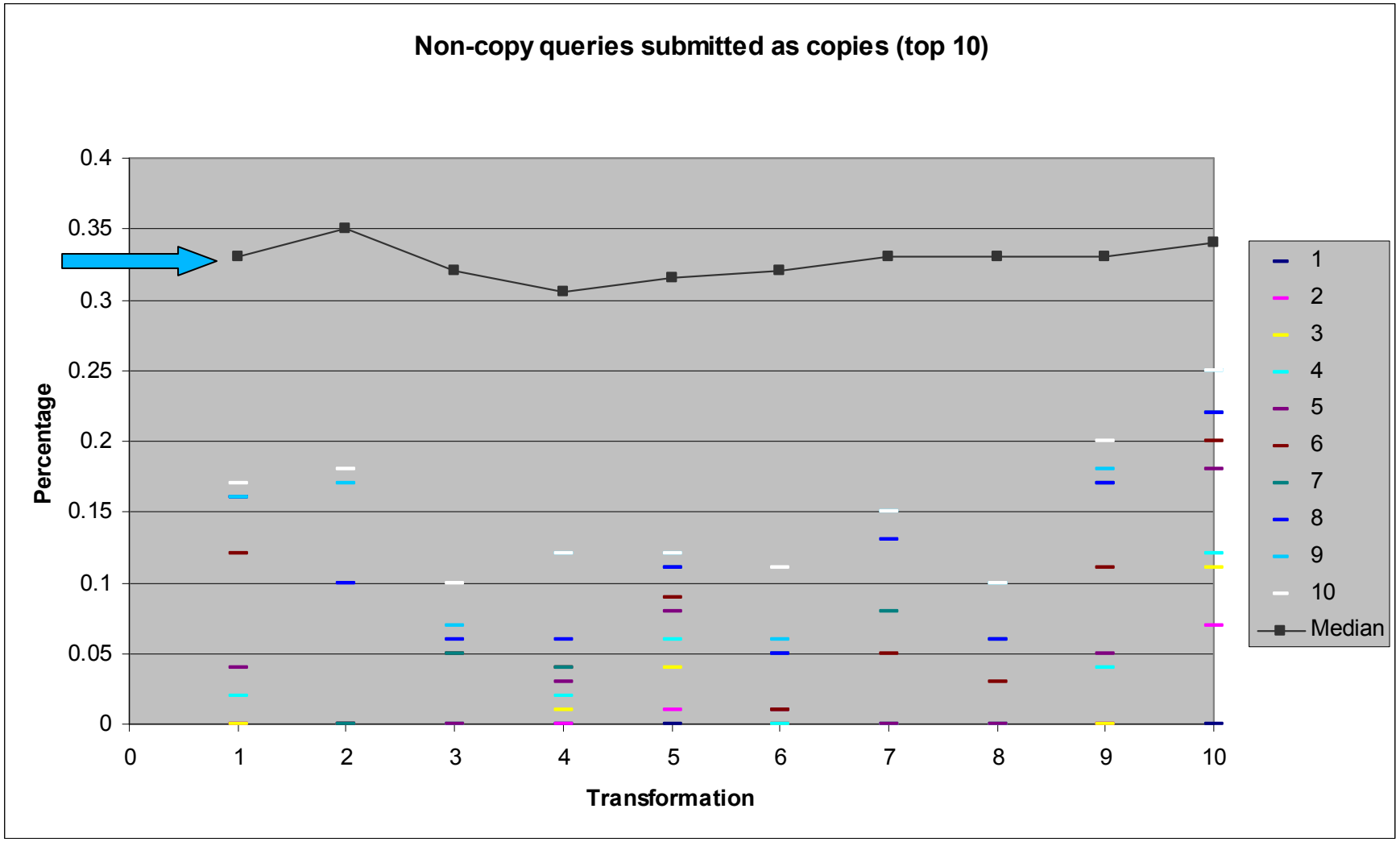
T5: Change of gamma

T6, T7: Decrease in quality

T8, T9: Post Production

T10: Random combination of 5 transformations

# CBCD evaluation (Top 10 performance)




T1: Cam Cording

T2: Pict. In Pict.

T3: Insertion of patterns

T4: Re-encoding

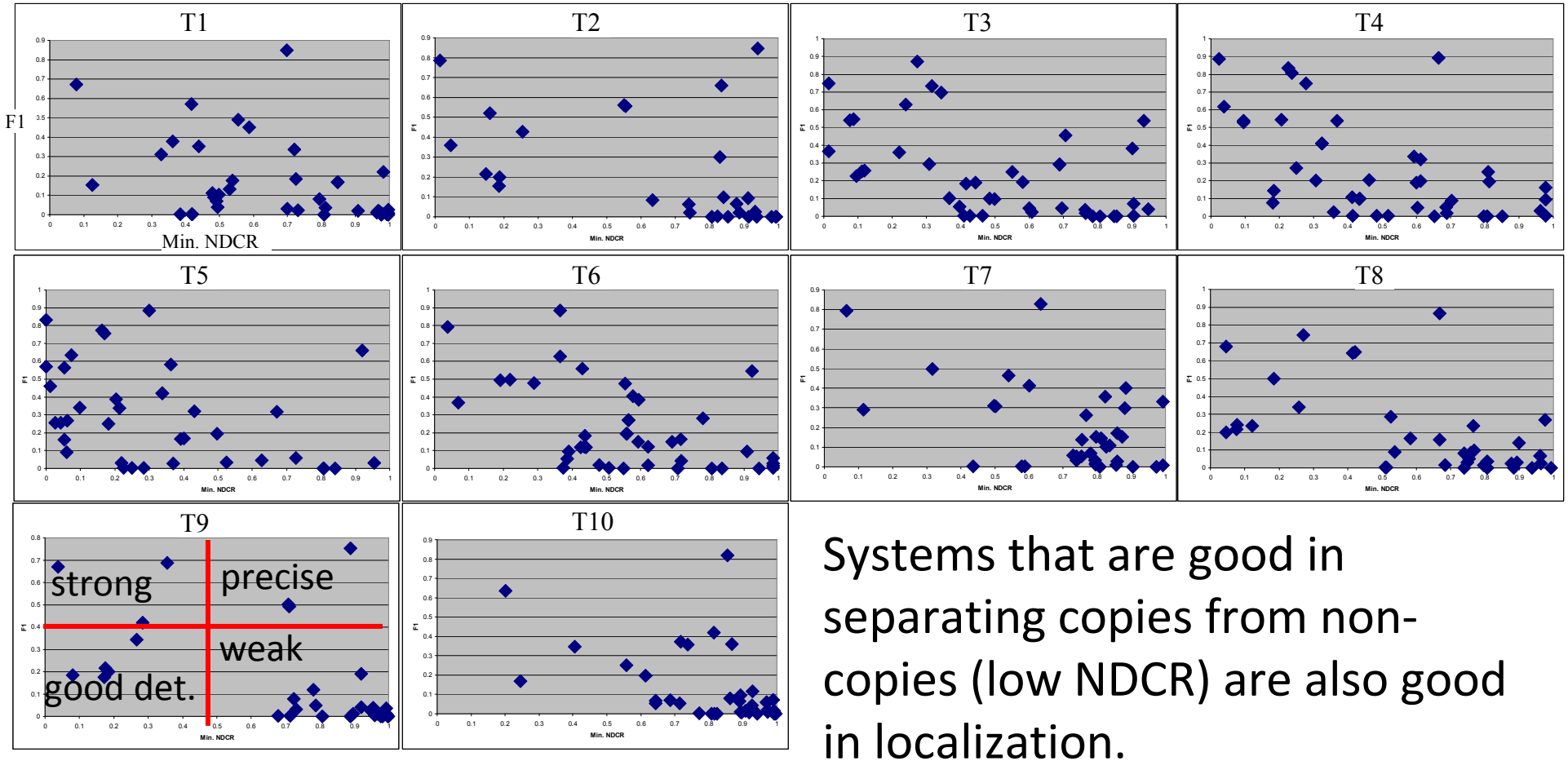
T5: Change of gamma

T6, T7: Decrease in quality

T8, T9: Post Production

T10: Random combination of 5 transformations

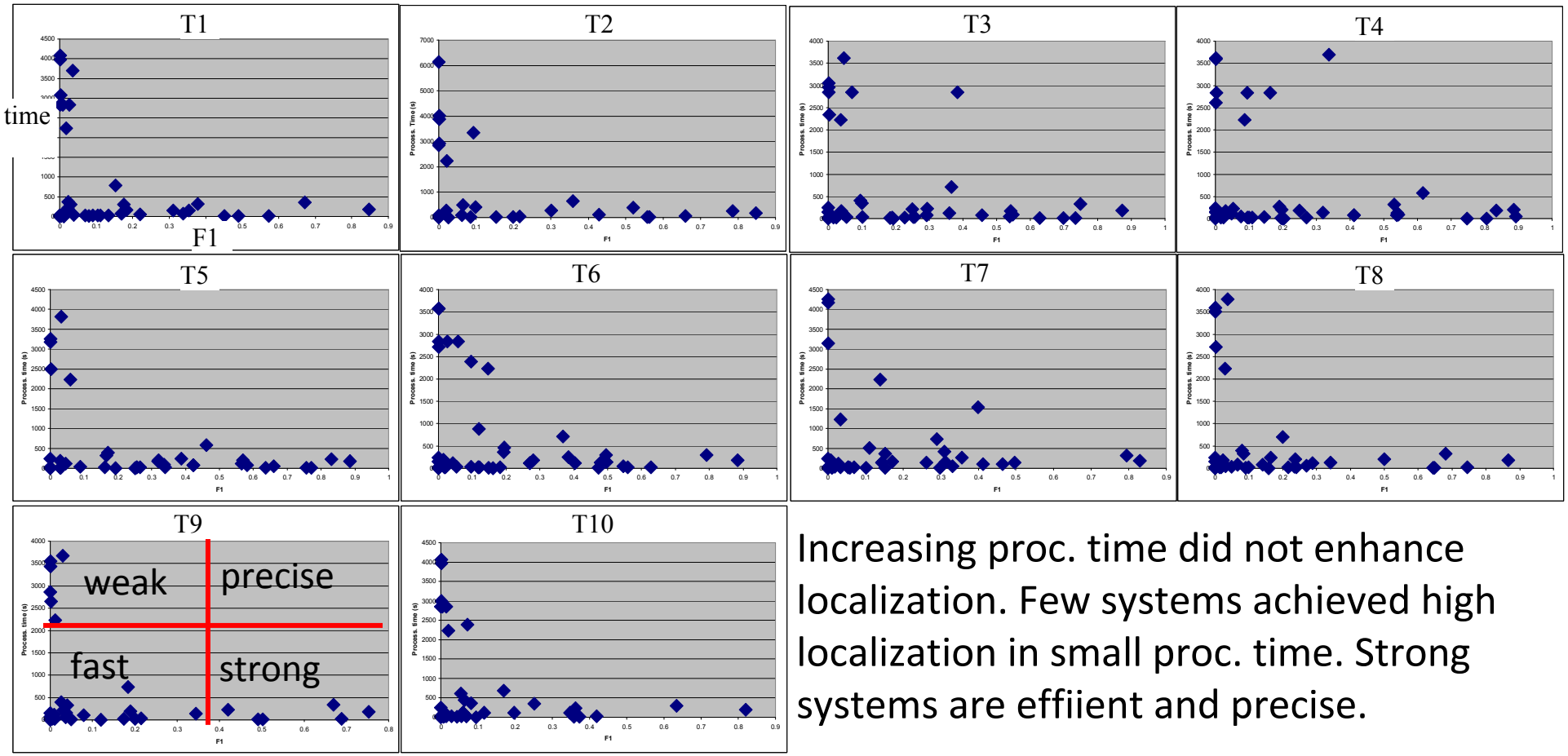
# Video-only performance (F1 vs Min. NDCR)




Systems that are good in separating copies from non-copies (low NDCR) are also good in localization.

T1: Cam Cording
T2: Pict. In Pict.
T3: Insertion of patterns
T4: Re-encoding
T5: Change of gamma
T6, T7: Decrease in quality
T8, T9: Post Production
T10: Random combination of 5 transformations

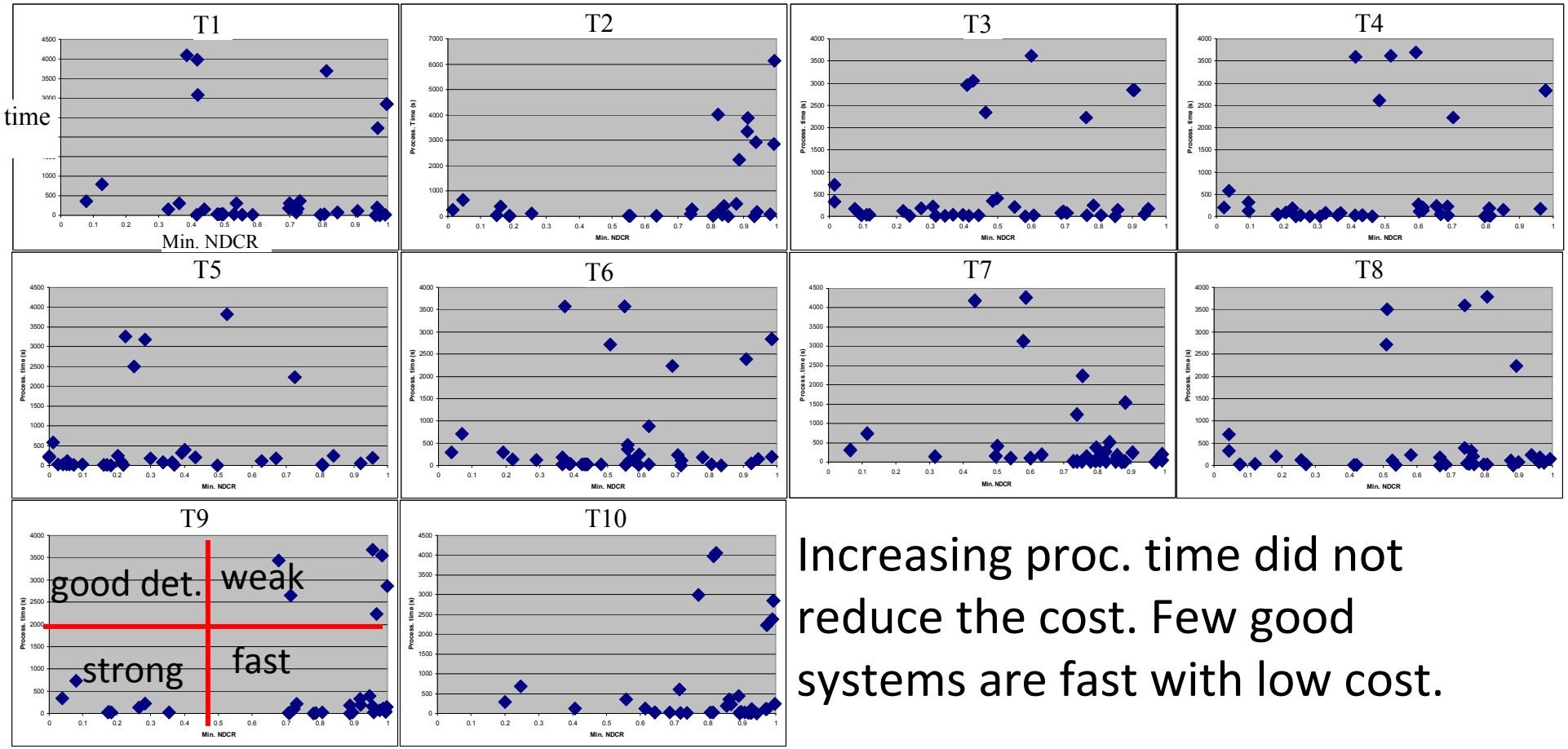
# Video-only perf. (Proc. time vs F1)




Increasing proc. time did not enhance localization. Few systems achieved high localization in small proc. time. Strong systems are effiient and precise.

T1: Cam Cording
T2: Pict. In Pict.
T3: Insertion of patterns
T4: Re-encoding
T5: Change of gamma
T6, T7: Decrease in quality
T8, T9: Post Production
T10: Random combination of 5 transformations

# Video-only perf. (Proc. time vs Min. NDCR)




Increasing proc. time did not reduce the cost. Few good systems are fast with low cost.

T1: Cam Cording
T2: Pict. In Pict.
T3: Insertion of patterns
T4: Re-encoding
T5: Change of gamma
T6, T7: Decrease in quality
T8, T9: Post Production
T10: Random combination of 5 transformations

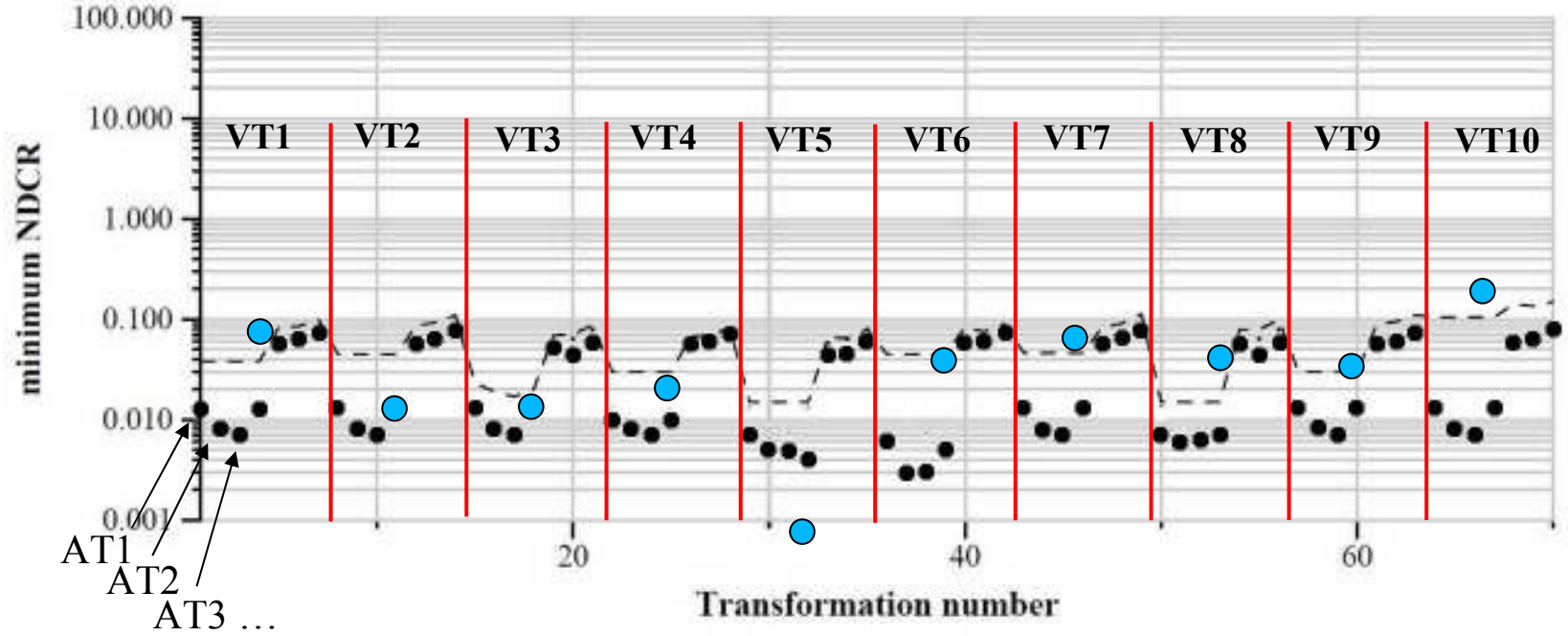
# Audio-only performance (1 run)

| | Min NDCR | F1 | Proc. Time |
|---|---|---|---|
| T1: nothing | 0.019 | 0.957 | 213.83 |
| T2: mp3 compression | 0.019 | 0.961 | 211.88 |
| T3: mp3 compression and multiband companding | 0.019 | 0.963 | 200.78 |
| T4: bandwidth limit and single-band companding | 0.019 | 0.962 | 200.93 |
| T5: mix with speech | 0.049 | 0.688 | 249.60 |
| T6: mix with speech, then multiband compress | 0.049 | 0.681 | 232.86 |
| T7: bandpass filter, mix with speech, compress | 0.064 | 0.605 | 236.58 |

One run submitted, its NDCR and F1 do rank among the best video only systems. Transformations 5,6 and 7 (variable mixing with unrelated audio content) seems to be harder.

# Audio+video runs

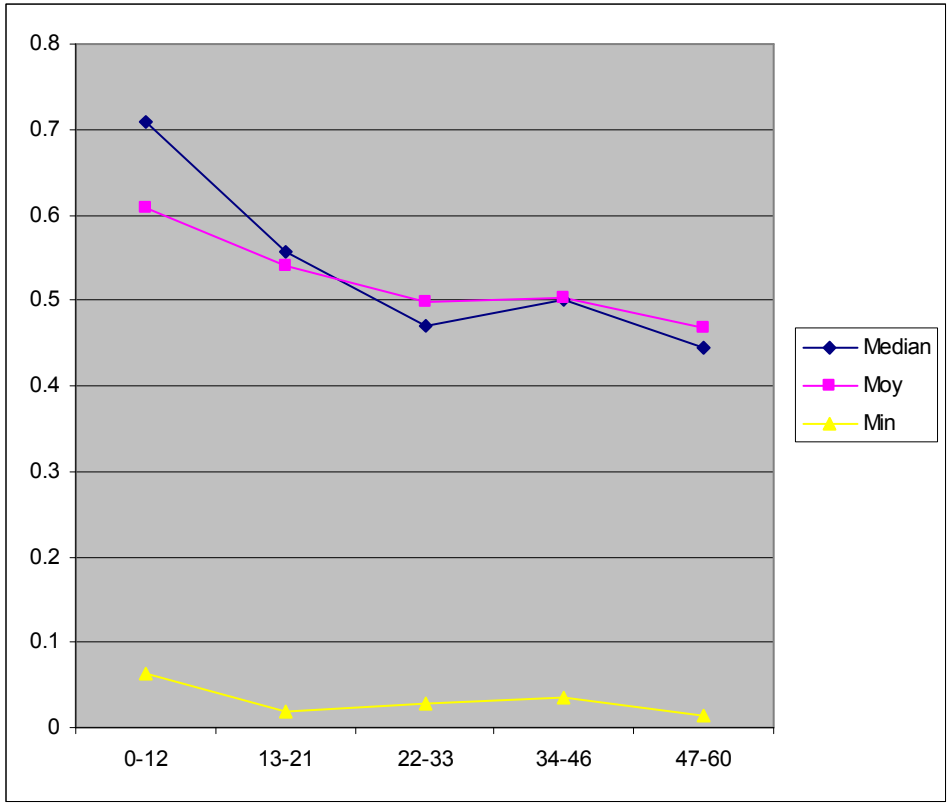
- ● – best video-only score for each V transformation
- ● – best video+audio score for each A+V transformation




- A+V run set too limited for broad conclusions, but …
- Relative effect of audio transformations seems similar across video transformations
- Using audio seems to help over just video when no speech is mixed in, except for VT5 (gamma)

# Query length impact

- Pmiss at min NDCR for 5 length intervals (Median, Avg and Min over all participants)




Slide contributed by Alexis Joly/IMEDIA

# Observations/questions

- All the pieces of the pilot came together as planned!
- Would not have been possible without major help from INRIA-IMEDIA, Laurent Joyeaux, Dan Ellis.
- Some systems have achieved very good results, the task has been difficult for many others.
- Score normalization across queries is critical
- Complex transformations are indeed more difficult. Query length has no major impact.
- Combination of a+v yields improvement
- How does the pilot task relate to a real operational CBCD task?

# IMEDIA: Availability of tools and test corpus

- 2 tools developed by INRIA IMEDIA team
  - **QueryComposer** (random composition of reference and non reference video materials)
  - **QueryTransformer** (about 20 transformations randomly parameterized and composable)
- Dissemination
  - Tools
    - **QueryComposer** binaries already available to TRECVID
    - **QueryTransformer** binaries available before end of 2008
    - Open sources available to everybody in 2009
  - Corpus
    - A copyright free corpus will be generated and made available to everybody in 2009 (probably based on MUSCLE/CIVR 2007 reference dataset)