TRECVID-2009 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
- 2009: first year after pilot task.
- Task has both a detection and localization component.
- Detection measure based on error rates.
- Weighted trade-off of type I and type II errors.
  - (false alarms vs. misses)
- Computation of optimal operating point by NIST.
- *Comparison of performance @ operating point submitted by participants (actual) with performance @ optimal operating point.*
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 1407 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

- For 2009 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 (required)

- At least 2 runs (for each of the two required tasks) are required representing two application profiles (“no false alarms”, “balanced”).

- Some groups submitted “video-only” runs but ignored the required “video + audio” task!!
INRIA query creation framework

**Step 1:** Selection of segments

**Step 2:** Transformations of segments

*figure 1: framework for building video queries*

**Hard cuts, mean length ref:** 32s, **mean length nonref:** 105s, **mean query length ~** 91 s
Datasets and queries

- **Dataset:**
  - Reference video collection:
    - Testing data: TV2009 (180 hr) and TV2007 and TV2008 S&V data (200 hr)
    - Development data: TV2007 and TV2008 S&V data (200 hr)
  - Non-reference video collection:
    - Testing data: TV2009 BBC rushes data (30 hrs)
    - Development data: TV2007 and TV2008 BBC rushes data (53 hrs)
- **Query types:** (Developed by INRIA-IMEDIA software run at NIST)
  - 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. 67 queries for each type.
- **After creating the queries, each was transformed.**
  - 7 video transformations by NIST (using a tool created by INRIA-IMEDIA)
  - 7 audio transformations by Dan Ellis at Columbia University
- **Yielding...** $7 \times 201 = 1407$ video queries, $7 \times 201 = 1407$ audio queries and $7 \times 7 \times 201 = 9849$ audio+video queries
Video transformations

- As requested in Tv2008, some transformations were not realistic and extreme (T7 and T9). This year 3 transformations were dropped:
  - T1 (camcording), T7 and T9.

- 7 Transformations were selected:
  - Picture in picture (T2)
  - Insertions of pattern (T3)
  - Strong re-encoding (T4)
  - Change of gamma (T5)
  - Frame dropping (T6)
  - Post production (T8) – by introducing 3 randomly selected combination of Crop, Shift, Contrast, Text insertion, Vertical mirroring, Insertion of pattern, Picture in picture,
  - Combination of 3 randomly selected transformations (T10) chosen from T2-T5, one transformation from Blur, Gamma, Frame dropping, Contrast, Compression, Ratio, White noise and T8.
Video transformations examples

- Picture in Picture
- Blur
- Insertion of pattern
- Strong re-encoding
- Noise
- Contrast
- Change in gamma
- Mirroring
- Ratio
- Crop
- Shift
- Text insertion
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 submit a run threshold, which defines the system’s operating point.
- Systems are asked to output a list of possible copies (each associated with a decision score).
- The run threshold is used to determine the asserted copies.
- A query can yield just one true positive.
- A query can give rise to many false alarms (even within one reference video).
- 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:
     - $P_{miss}$ (probability of a missed copy)
     - $R_{fa}$ (false alarm rate)
   - combines them using assumptions about two possible realistic scenarios:
     1. No False Alarm profile:
        - $Copy \ target \ rate \ (R_{target}) = 0.5/\text{hr}$
        - $Cost \ of \ a \ miss \ (C_{Miss}) = 1$
        - $Cost \ of \ a \ false \ alarm \ (C_{FA}) = 1000$
     2. Balanced profile:
        - $Copy \ target \ rate \ (R_{target}) = 0.5/\text{hr}$
        - $Cost \ of \ a \ miss \ (C_{Miss}) = 1$
        - $Cost \ of \ a \ false \ alarm \ (C_{FA}) = 1$

2. $F_1$ (how accurately the copy is located, harmonic mean of $P$ and $R$)
3. Mean processing time per query

Evaluation metrics (2)

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 a combination of F1-score and the decision score (using the Hungarian solution to the Bipartite Graph matching problem).

- The reference data has been found if and only if: The asserted test video ID is correct AND asserted copy and ref. video overlap.
Decision Error Tradeoff Curves $\text{Prob}_{\text{Miss}}$ vs. $\text{Rate}_{\text{FA}}$

Decision Score Histogram

Count of Observations

Decision Score

Full Distribution
Decision Error Tradeoff Curves $\text{Prob}_{\text{Miss}}$ vs. $\text{Rate}_{\text{FA}}$

Decision Score Histogram Separated wrt. Reference Annotation

- Incorrect System Observations
- True Observations
Decision Error Tradeoff Curves $\text{Prob}_{\text{Miss}}$ vs. $\text{Rate}_{\text{FA}}$

Decision Score Histogram Separated wrt. Reference Annotation $s$

$$P_{\text{Miss}}(\theta) = \frac{\# \text{MissedObs}}{\# \text{TrueObs}}$$

$$\text{Rate}_{\text{FA}}(\theta) = \frac{\# \text{FalseAlarms}}{\text{SignalDuration}}$$

signal: query
Decision Error Tradeoff Curves $\text{Prob}_{\text{Miss}}$ vs. $\text{Rate}_{\text{FA}}$

Compute $\text{Rate}_{\text{FA}}$ and $\text{P}_{\text{Miss}}$ for all $\Theta$

leads to: $\text{NDCR} = \text{P}_{\text{miss}} + \beta \text{ Rfa}$

The minimal $\beta$ defined by task characteristics

Optimal threshold determined by NIST
Example det curve: optimal vs actual NDCR
20 Participants (finishers) (2008: 22)

<table>
<thead>
<tr>
<th>Institution</th>
<th>ED</th>
<th>FE</th>
<th>SE</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asahikasei Co.</td>
<td>--</td>
<td>FE</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>AT&amp;T Labs - Research</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>Beijing University of Posts and Telecom.-MCPRL</td>
<td>ED</td>
<td>FE</td>
<td>SE</td>
<td>CD</td>
</tr>
<tr>
<td>Computer Research Institute of Montreal</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>Fudan University</td>
<td>--</td>
<td>FE</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>IBM Watson Research Center</td>
<td>ED</td>
<td>FE</td>
<td>SE</td>
<td>CD</td>
</tr>
<tr>
<td>Tsinghua University-IMG</td>
<td>ED</td>
<td>FE</td>
<td>SE</td>
<td>CD</td>
</tr>
<tr>
<td>Istanbul Technical University</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>JOANNEUM RESEARCH Forschungsgesellschaft mbH-JRS</td>
<td>ED</td>
<td>**</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>Chinese Academy of Sciences-MCG-ICT-CAS</td>
<td>--</td>
<td>--</td>
<td>SE</td>
<td>CD</td>
</tr>
<tr>
<td>Telefonica I+D</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>Tsinghua University-MPAM</td>
<td>--</td>
<td>FE</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>National Institute of Informatics</td>
<td>ED</td>
<td>FE</td>
<td>SE</td>
<td>CD</td>
</tr>
<tr>
<td>Nanjing University</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>TNO</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>TUBITAK UZAY</td>
<td>ED</td>
<td>FE</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>University of Brescia</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>City University of Hong Kong</td>
<td>ED</td>
<td>FE</td>
<td>SE</td>
<td>CD</td>
</tr>
<tr>
<td>University of Ottawa</td>
<td>ED</td>
<td>--</td>
<td>--</td>
<td>CD</td>
</tr>
<tr>
<td>Xi'an Jiaotong University</td>
<td>--</td>
<td>FE</td>
<td>SE</td>
<td>CD</td>
</tr>
</tbody>
</table>

-- : group didn’t participate, blue: new participant
## Submission types and counts

<table>
<thead>
<tr>
<th>Run type</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>V (video only)</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td>A (audio only)</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>M (video + audio)</td>
<td>6</td>
<td>42</td>
</tr>
<tr>
<td><strong>Total runs</strong></td>
<td>55</td>
<td>107</td>
</tr>
</tbody>
</table>

Good increase in a & m participation

<table>
<thead>
<tr>
<th>Video-only (balanced)</th>
<th>Video-only (nofa)</th>
<th>Audio-only (balanced)</th>
<th>Audio-only (nofa)</th>
<th>Video+Audio (balanced)</th>
<th>Video+Audio (nofa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>24</td>
<td>6</td>
<td>6</td>
<td>22</td>
<td>20</td>
</tr>
</tbody>
</table>

Balanced submissions between the two application profiles
Top “audio-only” runs

Act. Balanced

Act. NDCR

CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
CRIM.a,balanced,EnN22wr15
TNO.a,balanced.1
Top “audio-only” runs

Act. Nofa

Act. NDCR

T1
CRIM.a.nofa.EnN2pass
T1
CRIM.a.nofa.NN2para
T2
CRIM.a.nofa.NN2pass
T2
CRIM.a.nofa.EnN2pass
T3
CRIM.a.nofa.NN2pass
T3
CRIM.a.nofa.EnN2pass
T4
CRIM.a.nofa.NN2para
T4
CRIM.a.nofa.NN2para
T5
CRIM.a.nofa.EnN2pass
T5
CRIM.a.nofa.EnN2pass
T6
CRIM.a.nofa.NN2para
T6
CRIM.a.nofa.NN2para
T7
CRIM.a.nofa.EnN2pass
T7
CRIM.a.nofa.EnN2pass
T7
CRIM.a.nofa.EnN2pass
T7
TNO.a.nofa.1
Top “audio-only” runs

Opt. Balanced

Min. NDCR

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
Top “audio-only” runs

Opt. Nofa

Min. NDCR

0.1

0.2

0.3

0.4

0.5

0.6

0.7

0.8

0.9

1

T1

T1

T2

T2

T3

T3

T4

T4

T4

T4

T5

T5

T6

T6

T7

T7

CRIM.a.nofa.EnN2para

CRIM.a.nofa.EnN2para

CRIM.a.nofa.EnN2para

CRIM.a.nofa.EnN2para

CRIM.a.nofa.EnN2para

CRIM.a.nofa.EnN2para

TNO.a.nofa.1

CRIM.a.nofa.NN2para

CRIM.a.nofa.NN2para

CRIM.a.nofa.NN2para

CRIM.a.nofa.NN2para

CRIM.a.nofa.NN2para

CRIM.a.nofa.NN2para

CRIM.a.nofa.NN2para

CRIM.a.nofa.NN2para
Top “video-only” runs

Act. Balanced

<table>
<thead>
<tr>
<th>Act. NDCR</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T5</th>
<th>T8</th>
<th>T10</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCG-ICT-CAS.x,balanced.ICTBA.OPT</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCG-ICT-CAS.x,balanced.ICTBA.OPT</td>
<td></td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT11abs,x,balanced.12</td>
<td></td>
<td></td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCG-ICT-CAS.x,balanced.ICTBA.OPT</td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCG-ICT-CAS.x,balanced.ICTBA.OPT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLI-balanced.Embed01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>IBM-balanced.mixedBal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>
Top “video-only” runs

Act. Nofa

Act. NDCR

T2  MCG-ICT-CAS-MNOFA
    OPT

T3  MCG-ICT-CAS-MNOFA
    OPT

T4  ATT1abs-MNOFA

T5  ATT1lab-MNOFA

T6  TUBITAK-IZAYA-
    MNOFA

T8  ATT1ab-MNOFA

T10 ATT1ab-MNOFA
Top “video-only” runs

Opt. Nofa

Min. NDCR

T2  ATT Labs w.nofa.1
T3  ATT Labs w.nofa.1
T4  ATT Labs w.nofa.1
T5  ATT Labs w.nofa.1
T6  ATT Labs w.nofa.1
T8  ATT Labs w.nofa.1
T10 ATT Labs w.nofa.1
<table>
<thead>
<tr>
<th>Act. Nofa</th>
<th>CRIM.m.nofa.NN22para</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T2</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T3</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T4</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T5</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T6</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T7</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T8</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T9</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T10</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T11</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T12</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T13</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T14</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T15</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T16</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T17</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T18</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T19</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T20</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T21</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T22</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T23</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T24</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T25</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T26</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T27</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T28</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T29</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T30</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T31</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T32</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T33</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T34</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T35</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T36</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T37</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T38</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T39</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T40</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T41</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T42</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T43</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T44</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T45</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T46</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T47</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T48</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T49</td>
<td>CRIM.m.nofa.EnNN2pass</td>
</tr>
<tr>
<td>T50</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T51</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T52</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T53</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T54</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T55</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T56</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T57</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T58</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T59</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T60</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T61</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T62</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T63</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T64</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T65</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T66</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T67</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T68</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T69</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T70</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
<tr>
<td>T71</td>
<td>CRIM.m.nofa.NN22para</td>
</tr>
</tbody>
</table>

Top “video+audio” runs
Top “video + audio” runs

Opt. Balanced
CBCD video-only detection (Top 10 performance)

Balanced profile

Transformations

T6: Frame dropping  T8: Post Production  T10: Random combination of 3 transformations
CBCD audio-only detection (6 submitted runs)

Balanced profile

T1: nothing  T2: mp3 compression  T3: mp3 compression & multiband companding  T4: bandwidth limit & single-band companding
T5: mix with speech  T6: mix with speech, then multiband compress  T7: bandpass filter, mix with speech, compress
CBCD video+audio detection (Top 10 performance)
CBCD video-only localization (Top 10 performance)

Balanced profile

Transformations

- T3: Insertion of patterns
- T4: Strong Re-encoding
- T5: Change of gamma
- T6: Frame dropping
- T8: Post Production
- T10: Random combination of 3 transformations
CBCD audio-only localization (6 submitted runs)

Balanced profile

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>nothing</td>
</tr>
<tr>
<td>T2</td>
<td>mp3 compression</td>
</tr>
<tr>
<td>T3</td>
<td>mp3 compression &amp; multiband companding</td>
</tr>
<tr>
<td>T4</td>
<td>bandwidth limit &amp; single-band</td>
</tr>
<tr>
<td>T5</td>
<td>mix with speech</td>
</tr>
<tr>
<td>T6</td>
<td>mix with speech, then multiband compress</td>
</tr>
<tr>
<td>T7</td>
<td>bandpass filter, mix with speech, compress</td>
</tr>
</tbody>
</table>
CBCD video-only efficiency (Top 10 performance)

Balanced profile

T6: Frame dropping    T8: Post Production    T10: Random combination of 3 transformations
CBCD audio-only efficiency (6 submitted runs)

Balanced profile

Process. Time

Transformations

T1: nothing companding
T2: mp3 compression
T3: mp3 compression & multiband companding
T4: bandwidth limit & single-band compress
T5: mix with speech
T6: mix with speech, then multiband compress
T7: bandpass filter, mix with speech,
CBCD video+audio efficiency (Top 10 performance)

Balanced profile

Process. Time

Transformations

Median
1
2
3
4
5
6
7
8
9
10
CBCD video-only detection (Top 10 performance per T)

Nofa profile

optimal operating point is “reject all” for many systems

T6: Frame dropping  T8: Post Production  T10: Random combination of 3 transformations
CBCD audio-only detection (6 submitted runs)

Nofa profile

Min NDCR

Transformations

T1: nothing
T2: mp3 compression
T3: mp3 compression & multiband companding
T4: bandwidth limit & single-band compress
T5: mix with speech
T6: mix with speech, then multiband compress
T7: bandpass filter, mix with speech,
CBCD video+audio detection (Top 10 performance)
CBCD video-only localization (Top 10 performance)

Nofa profile

T6: Frame dropping  T8: Post Production  T10: Random combination of 3 transformations
CBCD audio-only localization (6 submitted runs)

Nofa profile

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

Transformations

F1
CBCD video-only efficiency (Top 10 performance)
CBCD audio-only efficiency (6 submitted runs)

Nofa profile

Transformations

1. nothing
2. mp3 compression
3. mp3 compression & multiband companding
4. bandwidth limit & single-band companding
5. mix with speech
6. mix with speech, then multiband compress
7. bandpass filter, mix with speech,
CBCD video+audio efficiency (Top 10 performance)

Nofa profile

Process Time

Error?lower than a or v

Transformations

Median

1

2

3

4

5

6

7

8

9

10
Increasing proc. time did not enhance localization. Few systems achieved high localization in small proc. time. Strong systems are efficient and precise.
Video only – **Balanced runs by transformations**

Most of the systems that are good in separating copies from non-copies (low NDCR) are also good in localization.
Video only – Balanced runs by transformations

Increasing proc. time did not reduce the cost. Few good systems are fast with low cost.
Most of the systems that are good in separating copies from non-copies (low NDCR) are also good in localization.
Increasing proc. time did not enhance localization. Few systems achieved high localization in small proc. time. Strong systems are efficient and precise.
Video only – Nofa runs by transformations

Increasing proc. time did not reduce the cost. Few good systems are fast with low cost.
Video+audio vs Video only (comparing best runs)

The m runs highly enhanced the detection accuracy across all transformations.
Video+audio vs Video only (comparing best runs)

The m runs helped in the majority of transformations to enhance localization
Comparing $a$, $v$, and $m$ best runs (Act. Balanced)
Comparing a, v, and m best runs (Act. Balanced)

Video (T6) - Balanced

Video (T8) - Balanced

Video (T10) - Balanced
Comparing a, v, and m best runs (Act. Nofa)
Comparing a, v, and m best runs (Act. Nofa)
Comparing $a$, $v$, and $m$ best runs (Opt. Balanced)
Comparing a, v, and m best runs (Opt. Balanced)
Comparing a, v, and m best runs (Opt. Nofa)
Comparing a, v, and m best runs (Opt. Nofa)

Video (T6) - Nofa

Video (T8) - Nofa

Video (T10) - Nofa

Audio transformations

Audio transformations

Audio transformations
## Lowest detection cost (NDCR) for individual transformations

<table>
<thead>
<tr>
<th></th>
<th>video only</th>
<th>audio only</th>
<th>mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>noFA profile</strong></td>
<td>MCG-ICT-CAS</td>
<td>CRIM</td>
<td>CRIM</td>
</tr>
<tr>
<td>ATT</td>
<td>ATT</td>
<td>TNO</td>
<td></td>
</tr>
<tr>
<td>NII</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Balanced profile</strong></td>
<td>MCG-ICT-CAS</td>
<td>CRIM</td>
<td>CRIM</td>
</tr>
<tr>
<td>ATT</td>
<td>ATT</td>
<td>TNO</td>
<td></td>
</tr>
<tr>
<td>TUBITAK</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Determining the optimal operating point

- New element for CBCD TV09, requires score normalization across queries
- For TV09 only some systems could do this
  - Large differences between actual and optimal results: big room for improvement
  - Huge impact on NDCR scores (esp. the video only runs)
  - Score normalization is critical
The influence of modalities

- Audio only detection results outperform video only
  - Easier? Techniques more mature?
- Combination of a+v improves upon a and v only
- Video only yields best localization results,
  - (still audio only systems have a higher median)
  - combination does not help
- Video only systems in general slightly faster
Comparison between noFA and balanced profiles

- tv08 discussion: teams are interested in a diversity of application profiles, noFA and balanced profiles were chosen for tv09
- Larger spread in NDCR for noFA profile (cost of a FA is high)
- noFA video only detection results slightly better than balanced
Trade-offs?

- The majority of low detection cost systems also have a good localization performance, but there is room for improvement here.
- TV9 data suggests no trade-off between detection cost and speed, and between localization and speed.
- Few systems perform well on all three measures.
Three evaluation measures

- The cost based NDCR evaluation measure seems suitable to model a variety of application profiles
  - large potential for improvement
- The localization and performance evaluation measures can help systems to find a balance in the accuracy/size/speed trade-off
  - top systems achieve near perfect results (F1>0.95)
- Only a minority of systems performs faster than RT
  - room for improvement
Other Observations

- Complex transformations are indeed more difficult.
- Limited attraction for audio-only queries.
- Many new teams, several strong tv08 teams did not participate this year.
- Would not have been possible without major help from INRIA-IMEDIA, Laurent Joyeaux, Dan Ellis.
Some trends in tv09 within site experiments

- Fusion of distinct frame representations (fingerprints)
  - SIFT descriptors
  - Block based features
  - Global (edge histogram)

- Speed optimization
  - GPU based local feature extraction

- Transformation detection + transformation specific approaches

- Score normalization
  - Dice coefficient, sigmoid transformation

- Combination of audio and video:
  - AND or OR
  - linear combination
Impact on real-world scenarios?

- How well do these results carry over to real application scenarios?
- Is the query creation process realistic?
  - copying audio track
  - hard cuts (no gradual transitions)
  - query lengths
- Do we have accurate estimates of Rfa and Pmiss?
- How realistic are the transformations?
- Transformations a-priori known
Some suggestions for a potential tv10 task

- single application profile
- retain three measures
- rethink query creation process
  - Need data for different scenarios
- Near similar detection?
Questions

- Did any one found multiple copies for a given query?
- Can we repeat the task again in Tv2010 on IA dataset?
- Any new thoughts about application profiles? Did the balanced/nofa achieved their goals?
- Any thoughts about “near similar” detection tasks?