Content-Based Video Copy Detection: PRISMA at TRECVID 2010

Juan Manuel Barrios and Benjamin Bustos

PRISMA Research Group
Department of Computer Science
University of Chile
{jbarrios,bebustos}@dcc.uchile.cl

November 17, 2010
Copy Detection System developed for TRECVID 2010.
Three Global descriptors.
No Audio information.
Pivot-based index with approximate search.
Voting algorithm for copy localization.
Implemented in C with OpenCV library.
System divided in five tasks/steps.
PRISMA System Overview

1. Preprocessing
2. Frame Sampling
3. Feature Extraction
4. Similarity Search
5. Copy Localization

Query Videos Reference Videos

Detection Result
System Tasks

1. **Preprocessing:**
   - Skip irrelevant frames.
   - Remove black borders.
   - Inverse transformations for Camcording, PIP and Flip.

Query videos increased from 1,608 to 5,378.
Reference videos kept in 11,524.
Frame Sampling:
- Divides each video in groups of similar consecutive frames (GF).
- Uniform subsampling of 3 frames per second.
- Similarity between frames defined as maximum difference between intensity of pixels.

Query Videos are divided into 1,000,000 groups.
Reference Videos are divided into 4,000,000 groups.
Frame Sampling:

- Divides each video in groups of similar consecutive frames (GF).
- Uniform subsampling of 3 frames per second.
- Similarity between frames defined as maximum difference between intensity of pixels.

Query Videos are divided into 1,000,000 groups.
Reference Videos are divided into 4,000,000 groups.
Feature Extraction:

- Descriptor of a group is the average of descriptors for each frame.
- Extracts three global visual descriptors:
  - EH: Edge Histogram \((4 \times 4 \times 10 = 160\) dimensions)
  - GH: Gray Histogram \((3 \times 3 \times 20 = 180\) dimensions)
  - CH: RGB Histogram \((2 \times 2 \times 48 = 192\) dimensions)

(1 byte per dimension)
Similarity Search:

- Compares descriptors from query groups with descriptors from reference groups.
- $DIST(G_i, G_j)$ is a distance function that measures the similarity between groups $G_i$ and $G_j$.
- $DIST$ is defined as a combination of two descriptors:
  - Run ehdNgyhst: $DIST$ combines EH and GH.
  - Run ehdNclrhst: $DIST$ combines EH and CH.
Similarity Search Task

- Distance between groups is a static weighted combination of distance between descriptors ($\gamma$):

$$\delta(G_i, G_j) = w_1 \times \gamma_1(G_i, G_j) + w_2 \times \gamma_2(G_i, G_j)$$

- We defined $\gamma$ as $L_1$ (Manhattan) distance for EHD, GH and CH vectors:

$$L_1(x, y) = \sum_{i=0}^{d} |x_i - y_i|$$

- Final distance between groups is the average of $\delta$ between three consecutive groups:

$$DIST(G_i, G_j) = \frac{\delta(G_{i-1}, G_{j-1}) + \delta(G_i, G_j) + \delta(G_{i+1}, G_{j+1})}{3}$$

- $DIST$ requires more than 1,000 operations to be evaluated.
We set weights for each descriptor using a histogram of distances between pairs of vectors.

Weights normalize to 100 the distance that covers 0.01% of pairs on each histogram: \( \frac{100}{1469} = 0.068 \quad \frac{100}{1106} = 0.090 \quad \frac{100}{660} = 0.152 \)

- **ehdNgryhst**: \( \delta = 0.068 \times EH + 0.090 \times GH \)
- **ehdNclrhst**: \( \delta = 0.068 \times EH + 0.152 \times CH \)
The intrinsic dimensionality $\frac{\mu^2}{2\sigma^2}$ quantifies how hard is to search on a metric space [Chávez et al, 2001].

Move $w_2$ to a value that locally maximizes intrinsic dimensionality of $\delta$.

Iterative algorithm that converged to:

- ehdNgryhst: $\delta = 0.068 \times EH + 0.090 \times GH$
- ehdNclrhst: $\delta = 0.068 \times EH + 0.045 \times CH$
Similarity Search Task

- The output of the Similarity Search task is a Nearest-Neighbors Table with most similar reference groups for each query group.

<table>
<thead>
<tr>
<th>Query</th>
<th>NN 1</th>
<th>NN 2</th>
<th>NN 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query1Group1</td>
<td>Vid07_Grp54</td>
<td>Vid08_Grp73</td>
<td>Vid01_Grp68</td>
</tr>
<tr>
<td>Query1Group2</td>
<td>Vid09_Grp13</td>
<td>Vid02_Grp34</td>
<td>Vid02_Grp33</td>
</tr>
<tr>
<td>Query1Group3</td>
<td>Vid07_Grp34</td>
<td>Vid03_Grp54</td>
<td>Vid09_Grp14</td>
</tr>
<tr>
<td>Query1Group4</td>
<td>Vid09_Grp15</td>
<td>Vid02_Grp13</td>
<td>Vid03_Grp65</td>
</tr>
<tr>
<td>Query1Group5</td>
<td>Vid01_Grp88</td>
<td>Vid01_Grp12</td>
<td>Vid07_Grp58</td>
</tr>
<tr>
<td>Query1Group6</td>
<td>Vid09_Grp54</td>
<td>Vid09_Grp17</td>
<td>Vid07_Grp59</td>
</tr>
<tr>
<td>Query1Group7</td>
<td>Vid01_Grp45</td>
<td>Vid03_Grp43</td>
<td>Vid03_Grp20</td>
</tr>
<tr>
<td>Query1Group8</td>
<td>Vid09_Grp19</td>
<td>Vid01_Grp12</td>
<td>Vid07_Grp61</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- A naive approach would evaluate $1,000,000 \times 4,000,000$ times $DIST$ (this takes about 11 month!).
Similarity Search Task

- $DIST$ complies with metric properties: Reflexivity, Non-Negativity, Symmetry, and Triangle Inequality.
- Let $q$ be a group of frames from a query video, and $v$ be a group of frames from a reference video.
- A lower bound for $DIST(q, v)$ can be calculated with pivots:
Similarity Search Task

- \( DIST \) complies with metric properties: Reflexivity, Non-Negativity, Symmetry, and Triangle Inequality.
- Let \( q \) be a group of frames from a query video, and \( v \) be a group of frames from a reference video.
- A lower bound for \( DIST(q, v) \) can be calculated with pivots:

\[
DIST(q, v) \geq |DIST(p, q) - DIST(p, v)|
\]
**Similarity Search Task**

- $DIST$ complies with metric properties: Reflexivity, Non-Negativity, Symmetry, and Triangle Inequality.
- Let $q$ be a group of frames from a query video, and $v$ be a group of frames from a reference video.
- A lower bound for $DIST(q, v)$ can be calculated with pivots:
  
  Let $S = \{p_1, ..., p_m\}$ be a set of pivots, then:
  \[
  DIST(q, v) \geq \max_{p \in S} \{|DIST(p, q) - DIST(p, v)|\}
  \]
Index creation:
- The system selects 4 sets of 9 pivots with the incremental SSS algorithm [Bustos et al, 2008].
  - Each set requires a table with $9 \times 4,000,000$ distances.
- The system compares the 4 sets and selects the set that has the greatest average lower bound and discards the others [Zezula et al, 2005].
Index creation:
- The system selects 4 sets of 9 pivots with the incremental SSS algorithm [Bustos et al, 2008].
  - Each set requires a table with $9 \times 4,000,000$ distances.
- The system compares the 4 sets and selects the set that has the greatest average lower bound and discards the others [Zezula et al, 2005].
Similarity Search Task

- Similarity search for a query group \( q \):
  - For every pivot \( p \) evaluate \( DIST(q, p) \).
  - For every reference group \( v \) calculate a lower bound for \( DIST(q, v) \).
    - Only 9 operations to calculate each lower bound.
  - Select 4,000 objects (0.1\%) with lowest lower bounds.
  - Calculate actual \( DIST(q, v) \) just for the 4,000 objects and select the NNs between them.
Similarity Search Task

- Similarity search for a query group $q$:
  - For every pivot $p$ evaluate $DIST(q, p)$.
  - For every reference group $v$ calculate a lower bound for $DIST(q, v)$
    - Only 9 operations to calculate each lower bound.
  - Select 4,000 objects (0.1%) with lowest lower bounds.
  - Calculate actual $DIST(q, v)$ just for the 4,000 objects and select the NNs between them.
Similarity Search Task

- Similarity search for a query group $q$:
  - For every pivot $p$ evaluate $DIST(q, p)$.
  - For every reference group $v$ calculate a lower bound for $DIST(q, v)$
    - Only 9 operations to calculate each lower bound.
  - Select 4,000 objects (0.1%) with lowest lower bounds.
  - Calculate actual $DIST(q, v)$ just for the 4,000 objects and select the NNs between them.
Similarity Search Task

- Similarity search for a query group \( q \):
  - For every pivot \( p \) evaluate \( DIST(q, p) \).
  - For every reference group \( v \) calculate a lower bound for \( DIST(q, v) \):
    - Only 9 operations to calculate each lower bound.
  - Select 4,000 objects (0.1\%) with lowest lower bounds.
  - Calculate actual \( DIST(q, v) \) just for the 4,000 objects and select the NNs between them.
Similarity Search Task

- Similarity search for a query group $q$:
  - For every pivot $p$ evaluate $DIST(q, p)$.
  - For every reference group $v$ calculate a lower bound for $DIST(q, v)$
    - Only 9 operations to calculate each lower bound.
  - Select 4,000 objects (0.1%) with lowest lower bounds.
  - Calculate actual $DIST(q, v)$ just for the 4,000 objects and select the NNs between them.
Similarity Search Task

- Similarity search for a query group $q$:
  - For every pivot $p$ evaluate $DIST(q, p)$.
  - For every reference group $v$ calculate a lower bound for $DIST(q, v)$
    - Only 9 operations to calculate each lower bound.
  - Select 4,000 objects (0.1%) with lowest lower bounds.
  - Calculate actual $DIST(q, v)$ just for the 4,000 objects and select the NNs between them.
Copy Localization:

- Takes NNs table and searches for chains of groups belonging to a same reference video with temporal coherence.
- Voting algorithm based on NN rank, NN distance and spread of votes in chain.
- Copy localization set as start/end of chain.

<table>
<thead>
<tr>
<th>Query</th>
<th>NN 1</th>
<th>NN 2</th>
<th>NN 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query1Group1</td>
<td>Vid07_Grp54</td>
<td>Vid08_Grp73</td>
<td>Vid01_Grp68</td>
</tr>
<tr>
<td>Query1Group2</td>
<td>Vid09_Grp13</td>
<td>Vid02_Grp34</td>
<td>Vid02_Grp33</td>
</tr>
<tr>
<td>Query1Group3</td>
<td>Vid07_Grp34</td>
<td>Vid03_Grp54</td>
<td>Vid09_Grp14</td>
</tr>
<tr>
<td>Query1Group4</td>
<td>Vid09_Grp15</td>
<td>Vid02_Grp13</td>
<td>Vid03_Grp65</td>
</tr>
<tr>
<td>Query1Group5</td>
<td>Vid01_Grp88</td>
<td>Vid01_Grp12</td>
<td>Vid07_Grp58</td>
</tr>
<tr>
<td>Query1Group6</td>
<td>Vid09_Grp54</td>
<td>Vid09_Grp17</td>
<td>Vid07_Grp59</td>
</tr>
<tr>
<td>Query1Group7</td>
<td>Vid01_Grp45</td>
<td>Vid03_Grp43</td>
<td>Vid03_Grp20</td>
</tr>
<tr>
<td>Query1Group8</td>
<td>Vid09_Grp19</td>
<td>Vid01_Grp12</td>
<td>Vid07_Grp61</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Copy Localization:

- Takes NNs table and searches for chains of groups belonging to a same reference video with temporal coherence.
- Voting algorithm based on NN rank, NN distance and spread of votes in chain.
- Copy localization set as start/end of chain.

<table>
<thead>
<tr>
<th>Query</th>
<th>NN 1</th>
<th>NN 2</th>
<th>NN 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query1Group1</td>
<td>Vid07_Grp54</td>
<td>Vid08_Grp73</td>
<td>Vid01_Grp68</td>
</tr>
<tr>
<td>Query1Group2</td>
<td>Vid09_Grp13</td>
<td>Vid02_Grp34</td>
<td>Vid02_Grp33</td>
</tr>
<tr>
<td>Query1Group3</td>
<td>Vid07_Grp34</td>
<td>Vid03_Grp54</td>
<td>Vid09_Grp14</td>
</tr>
<tr>
<td>Query1Group4</td>
<td>Vid09_Grp15</td>
<td>Vid02_Grp13</td>
<td>Vid03_Grp65</td>
</tr>
</tbody>
</table>
| Query1Group5   | Vid01_Grp88   | Vid01_Grp12   | **Vid07_Grp58** | dist
| Query1Group6   | Vid09_Grp54   | Vid09_Grp17   | **Vid07_Grp59** | dist
| Query1Group7   | Vid01_Grp45   | Vid03_Grp43   | Vid03_Grp20   |
| Query1Group8   | Vid09_Grp19   | Vid01_Grp12   | **Vid07_Grp61** | dist
| ...            | ...           | ...           | ...           |

score Vid07 = 2.2
5 Copy Localization:

- Takes NNs table and searches for chains of groups belonging to a same reference video with temporal coherence.
- Voting algorithm based on NN rank, NN distance and spread of votes in chain.
- Copy localization set as start/end of chain.

<table>
<thead>
<tr>
<th>Query</th>
<th>NN 1</th>
<th>NN 2</th>
<th>NN 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query1Group1</td>
<td>Vid07_Grp54 dist</td>
<td>Vid08_Grp73 dist</td>
<td>Vid01_Grp68 dist</td>
</tr>
<tr>
<td>Query1Group2</td>
<td>Vid09_Grp13 dist</td>
<td>Vid02_Grp34 dist</td>
<td>Vid02_Grp33 dist</td>
</tr>
<tr>
<td>Query1Group3</td>
<td>Vid07_Grp34 dist</td>
<td>Vid03_Grp54 dist</td>
<td>Vid03_Grp65 dist</td>
</tr>
<tr>
<td>Query1Group4</td>
<td>Vid09_Grp15 dist</td>
<td>Vid02_Grp13 dist</td>
<td>Vid07_Grp58 dist</td>
</tr>
<tr>
<td>Query1Group5</td>
<td>Vid01_Grp88 dist</td>
<td>Vid01_Grp12 dist</td>
<td>Vid07_Grp59 dist</td>
</tr>
<tr>
<td>Query1Group6</td>
<td>Vid09_Grp54 dist</td>
<td>Vid03_Grp43 dist</td>
<td>Vid03_Grp20 dist</td>
</tr>
<tr>
<td>Query1Group7</td>
<td>Vid01_Grp45 dist</td>
<td>Vid01_Grp12 dist</td>
<td>Vid07_Grp61 dist</td>
</tr>
<tr>
<td>Query1Group8</td>
<td>Vid09_Grp19 dist</td>
<td>Vid01_Grp61 dist</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- score Vid09 = 3.7
- score Vid07 = 2.2
Results

- Submitted Runs:
  - balanced.ehdNgryhst: $\delta = 0.068 \times EH + 0.090 \times GH$
  - balanced.ehdNclrhst: $\delta = 0.068 \times EH + 0.045 \times CH$
  - nofa.ehdNgryhst: equal to balanced.ehdNgryhst with stricter voting algorithm.
  - nofa.ehdNghT10: equal to nofa.ehdNgryhst but with a different threshold.

- Analysis focused on Optimal NDCR.
- EH+GH slightly better than EH+CH.
- Better results in NOFA profile than in Balanced profile.
Optimal NDCR:
- Lower NDCR than median for each transformation.
- Better results for Insertion of Pattern and Strong Reencoding.
Optimal NDCR:
- Lower NDCR than median for each transformation.
- Better results for Insertion of Pattern and Strong Reencoding.
• Optimal F1:
  • Good localization for PIP and bad localization for Camcording and Change in gamma.
• Mean Time:
  • Slightly higher than the median, specially for camcording and PIP.
Comparison

- Comparison with Optimal NDCR averaged between all transformations.
- 22 teams, 41 submitted runs for balanced profile and 37 for nofa profile.

<table>
<thead>
<tr>
<th>Run</th>
<th>Avg Opt NDCR</th>
<th>global rank</th>
<th>video-only rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>balanced.ehdNgryhst</td>
<td>0.597</td>
<td>14\textsuperscript{th} of 41</td>
<td>1\textsuperscript{st} of 15</td>
</tr>
<tr>
<td>balanced.ehdNclrhst</td>
<td>0.658</td>
<td>16\textsuperscript{th} of 41</td>
<td>3\textsuperscript{rd} of 15</td>
</tr>
<tr>
<td>nofa.ehdNgryhst</td>
<td>0.611</td>
<td>10\textsuperscript{th} of 37</td>
<td>1\textsuperscript{st} of 14</td>
</tr>
<tr>
<td>nofa.ehdNghT10</td>
<td>0.611</td>
<td>11\textsuperscript{th} of 37</td>
<td>2\textsuperscript{nd} of 14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Run</th>
<th>Avg Opt F1</th>
<th>global rank</th>
<th>video-only rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>balanced.ehdNgryhst</td>
<td>0.820</td>
<td>15\textsuperscript{th} of 41</td>
<td>2\textsuperscript{nd} of 15</td>
</tr>
<tr>
<td>balanced.ehdNclrhst</td>
<td>0.820</td>
<td>16\textsuperscript{th} of 41</td>
<td>3\textsuperscript{rd} of 15</td>
</tr>
<tr>
<td>nofa.ehdNgryhst</td>
<td>0.828</td>
<td>14\textsuperscript{th} of 37</td>
<td>1\textsuperscript{st} of 14</td>
</tr>
<tr>
<td>nofa.ehdNghT10</td>
<td>0.828</td>
<td>15\textsuperscript{th} of 37</td>
<td>2\textsuperscript{nd} of 14</td>
</tr>
</tbody>
</table>
Comparison

No False Alarms Profile

Average Optimal F1

Average Optimal NDCR

PRISMA

Video only
Audio only
Audio+Video

No False Alarms Profile (logarithmic scale)

PRISMA (University of Chile)  CCD Task  November 17, 2010  22 / 25
Comparison

Balanced Profile

- Video only
- Audio only
- Audio+Video

PRISMA

PRISMA (University of Chile)
Conclusions

- Acceptable overall results:
  - Global descriptors can achieve competitive results with TRECVID transformations.
  - Pivot-based approximation enables to discard 99.9% of distance computations and still have good effectiveness.

- Two novel techniques:
  - Set weights maximizing intrinsic dimensionality.
  - Calculate actual distance just for 0.1% lowest lower bounds.

- Future work:
  - Improve the efficiency of preprocessing task.
  - Test other distances for descriptors instead of $L_1$ (in particular some non-metric similarity measure).
  - Test the inclusion of audio information and local descriptors.
Thank you!

Juan Manuel Barrios and Benjamin Bustos

PRISMA Research Group
Department of Computer Science
University of Chile
{jbarrios,bebustos}@dcc.uchile.cl

November 17, 2010

Thank you!