CDVP & TRECVID-2003
News Story Segmentation Task

Csaba Czirjek, Gareth J.F. Jones, Seán Marlow, Noel Murphy, Noel E. O’Connor, Neil O’Hare, Alan F. Smeaton
Contents

• Introduction
  – Structure of News Broadcast
  – System Overview

• Story Segmentation System
  – Feature Extraction Process
  – Combination of Features using Support Vector Machine
  – Submitted Runs

• Results

• Conclusions
Structure of a News Broadcast

• We assume stories are delimited by shots of the anchorperson

• Features of Anchor shots:
  – All anchor shots within a broadcast taken from the same camera setup
  – filmed with a static camera, with little object motion
  – anchor shots in a single broadcast are visually similar to each other
Structure of a News Broadcast

- Anchorperson Shots
- Commercial Break
- News Report Shots

Diagram showing the structure of a news broadcast with various segments and shots.
System Overview

- We use TRECVID 2003 common shot boundary provided by CLIPS-IMAG
- Extracted features combined to detect anchor shots
- Story boundaries logged at the start of anchor shots
- Aim is to extract features that are robust to changes across broadcasters (e.g., faces, motion, shot length)
- This would give a generic news segmentation system
System Overview

Shot Level Feature Extraction
- Shot Clustering
- Face Detection
- Motion Activity Analysis x 2
- Shot Length
- Text Segmentation

News Story Detection
- Support Vector Machine

News Stories
1
2
3
4
5
6
7
8

30 Minute News Program
Shot Boundary Detection
Donated by CLIPS-IMAG

Donated by StreamSage
• Shots are clustered based on visual similarity (colour histogram)

• anchor shots grouped together

• anchor clusters identified using heuristics:
  – tend to be dispersed throughout the broadcast
  – average length longer than others
  – anchor shots are very similar to each other: they form ‘tighter’ clusters
Feature Extraction 2 - Face Detection

- Coarse to fine approach to extract candidate regions:
  - Skin like pixels identified based on colour
  - Morphological filtering used to obtain smoothed areas of connected pixels
  - Shape and size heuristics remove candidate face regions

- Candidates passed to a Principle Component Analysis (PCA) module for final classification

- Every 12th frame (I-frames) used for processing
Face Detection

Original video file

For every 12th frame

Filtered image after morphological adjustment

Skin filtering + morphological adjustment

Size/shape heuristics

Image after applying size/shape heuristics

Detected faces with confidence score

Face Database

PCA
Feature Extraction 3 - Activity Measure

- Motion Activity analysis based on MPEG-1 motion vectors
- Every P-frame is analysed
- We count the number of zero length motion vectors in a P-frame (excluding I-blocks)
- Activity measure:

\[
\frac{\text{No. of zero length vectors}}{\text{Total No. of macroblocks}}
\]
Two separate shot level measures used:

- least active P-frame is used to represent the shot
- All motion vectors across a shot are added to form a cumulative motion vector. Activity measure then calculated using cumulative motion vector

\[
\begin{align*}
\text{frame } a & & \text{frame } b & & \text{cumulative frame:} \\
0, -1 & 0,1 & -3,5 & 0,1 & 1,0 & -2,4 & 0,0 & 1,1 & -5,9 \\
0,0 & 0,0 & 4,3 & 3,0 & 0,0 & 0,0 & 3,0 & 0,0 & 4,3 \\
-2,1 & 1,-1 & 1,0 & -2,1 & 0,1 & 0,1 & -4,2 & 1,0 & 1,1
\end{align*}
\]
• Shot length used as a feature

• Measured in frames
• To allow us to complete the required runs, we used text analysis provided by StreamSage

• StreamSage text output used as binary feature
Combination of Features - SVM

- Extracted features combined using Support Vector Machine
- Trained on 10 hours of the TRECVID 2003 development set (5 CNN, 5 ABC)
- Resulting SVM classifier detects anchor shots
- Story boundaries are logged at the beginning of anchor shots
Submitted Runs

• 3 Required Runs
  - A/V only system - generic system for ABC and CNN (DCU03_REQ_AV)
  - A/V + text - generic system for ABC and CNN (DCU03_REQ_AV_TEXT)
  - Text only - text Analysis provided by StreamSage (DCU03_REQ_TEXT_ONLY)

• 2 Additional Optional Runs
  - Specialised systems for ABC and CNN. Separate SVMs for each broadcaster (DCU03_OPT_AV)
  - Clustering algorithm in isolation (DCU03_OPT_CLUSTER)
## DCU Results

<table>
<thead>
<tr>
<th>System ID</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCU03_REQ_AV</td>
<td>0.328</td>
<td>0.409</td>
</tr>
<tr>
<td>DCU03_REQ_AV_TEXT</td>
<td>0.294</td>
<td>0.453</td>
</tr>
<tr>
<td>DCU03_REQ_TEXT_ONLY</td>
<td>0.049</td>
<td>0.208</td>
</tr>
<tr>
<td><strong>DCU03_OPT_AV</strong></td>
<td><strong>0.313</strong></td>
<td><strong>0.453</strong></td>
</tr>
<tr>
<td>DCU03_OPT_CLUSTER</td>
<td>0.364</td>
<td>0.304</td>
</tr>
</tbody>
</table>
Conclusions

- Best results from specialised system (DCU03_OPT_AV)
- Generic system not far behind
- Extracted features robust across broadcasters
- Combined results improve precision with small loss in recall compared to clustering alone