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

CENTER FOR DIGITAL VIDEO PROCESSING

- Introduction
  - Structure of News Broadcast
  - System Overview
- Story Segmentation System
  - Feature Extraction Process
  - Combination of Features using Support Vector Machine
  - Submitted Runs
- Results
- Conclusions

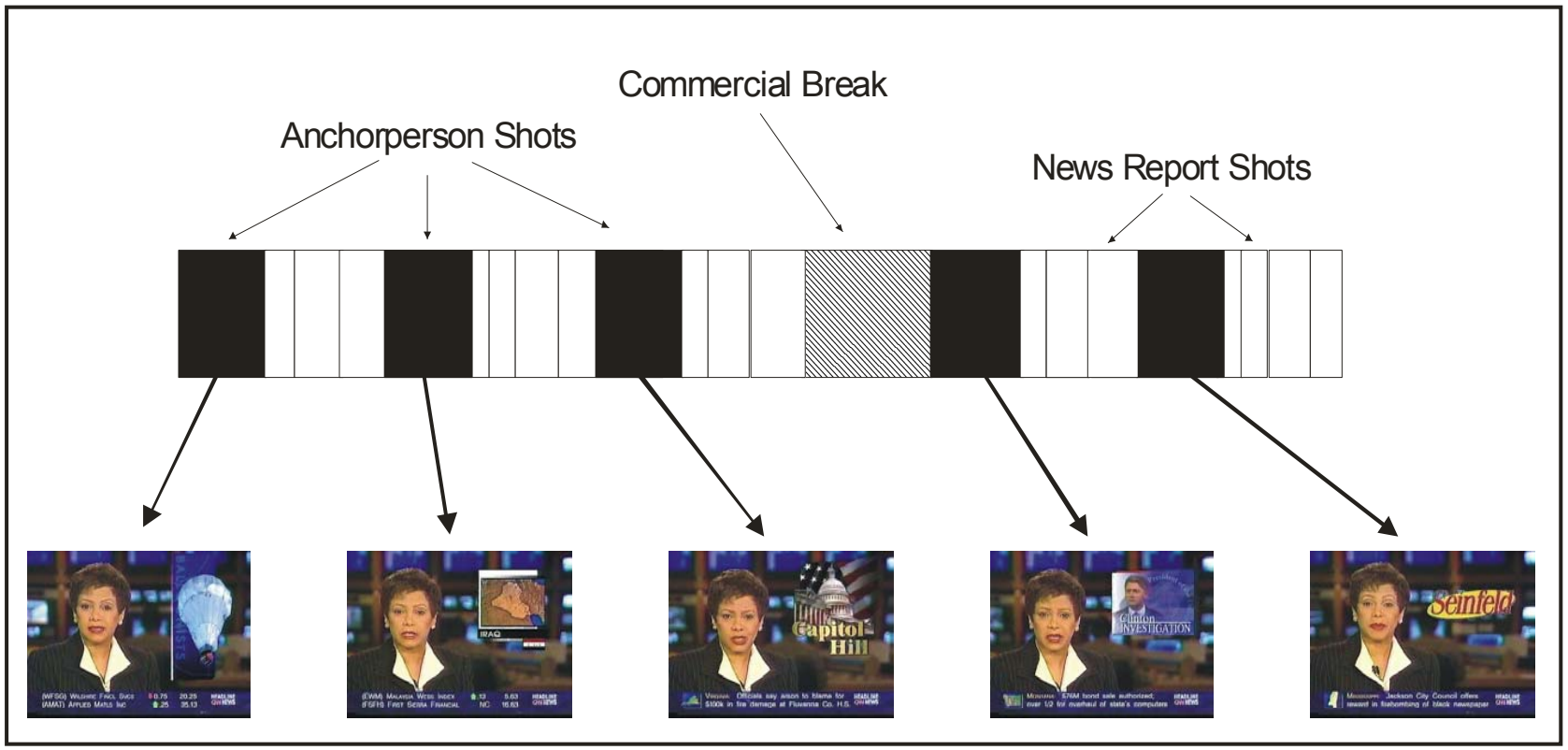
Structure of a News Broadcast

- 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

CENTER FOR DIGITAL VIDEO PROCESSING



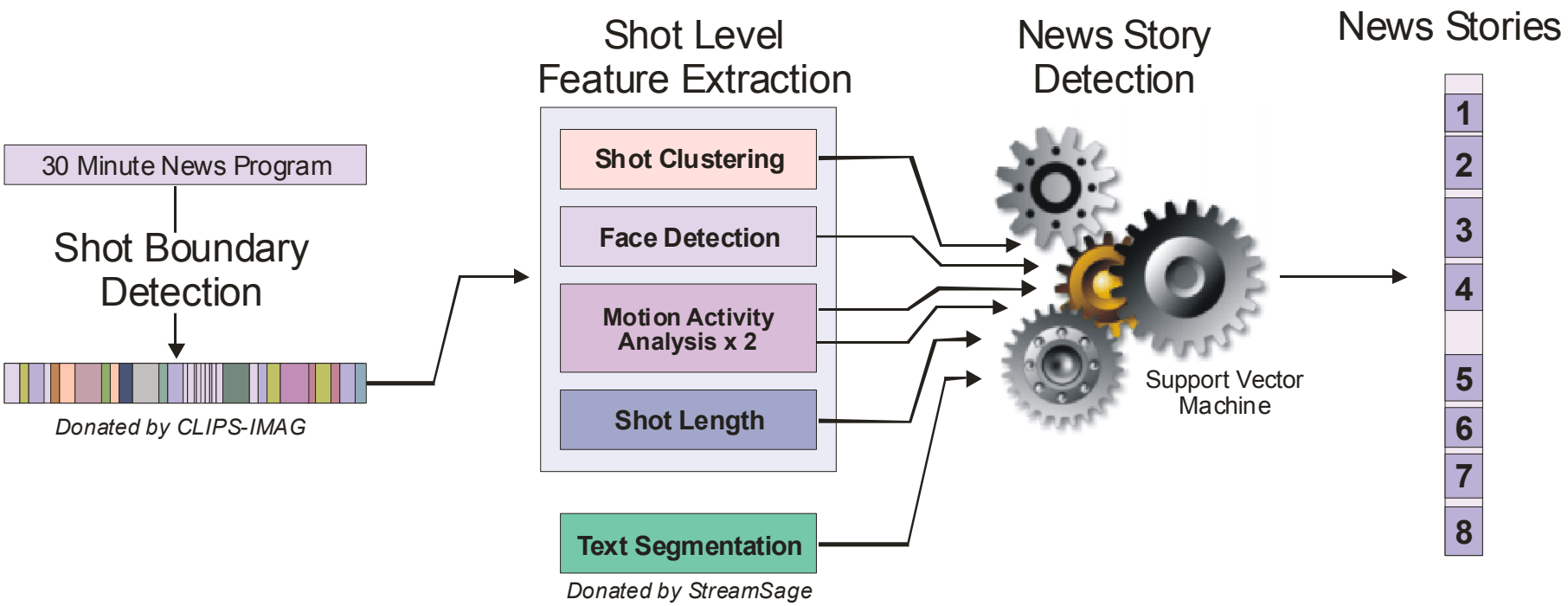
System Overview

CENTER FOR DIGITAL VIDEO PROCESSING

- 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
  (eg faces, motion, shot length)
- This would give a generic news segmentation system

System Overview

CENTER FOR DIGITAL VIDEO PROCESSING



Feature Extraction 1 - Shot Clustering

CENTER FOR DIGITAL VIDEO PROCESSING

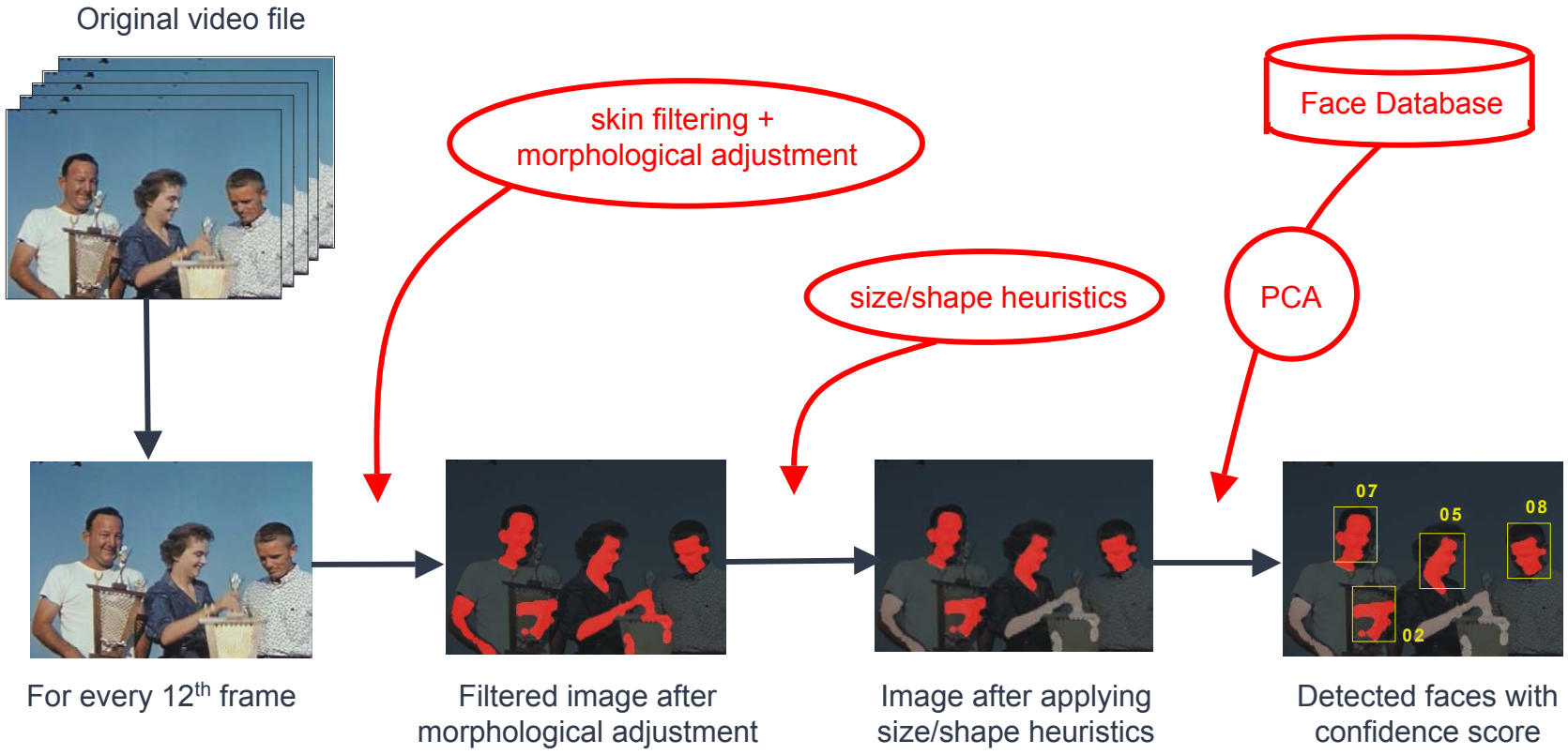
- Shots are clustered based on visual similarity (colour histogram)
- anchor shots grouped together
- anchor clusters identified using heuristics:
  - tend to be dispersed throughout the broadcast
  - average length longer than others
  - anchor shots are **very** similar to each other: they form 'tighter' clusters

Feature Extraction 2 - Face Detection

CENTER FOR DIGITAL VIDEO PROCESSING

- 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



Feature Extraction 3 - Activity Measure

CENTER FOR DIGITAL VIDEO PROCESSING

- 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:

$$\frac{\text{No. of zero length vectors}}{\text{Total No. of macroblocks}}$$

Feature Extraction 3 - Activity Measure

CENTER FOR DIGITAL VIDEO PROCESSING

- 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

frame a				frame b			=	cumulative frame: frame a + frame b		
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

Feature Extraction 4 - Shot Length

- Shot length used as a feature
- Measured in frames

Feature Extraction 5 - Text Analysis

- 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

CENTER FOR DIGITAL VIDEO PROCESSING

- 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

CENTER FOR DIGITAL VIDEO PROCESSING

- 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)**

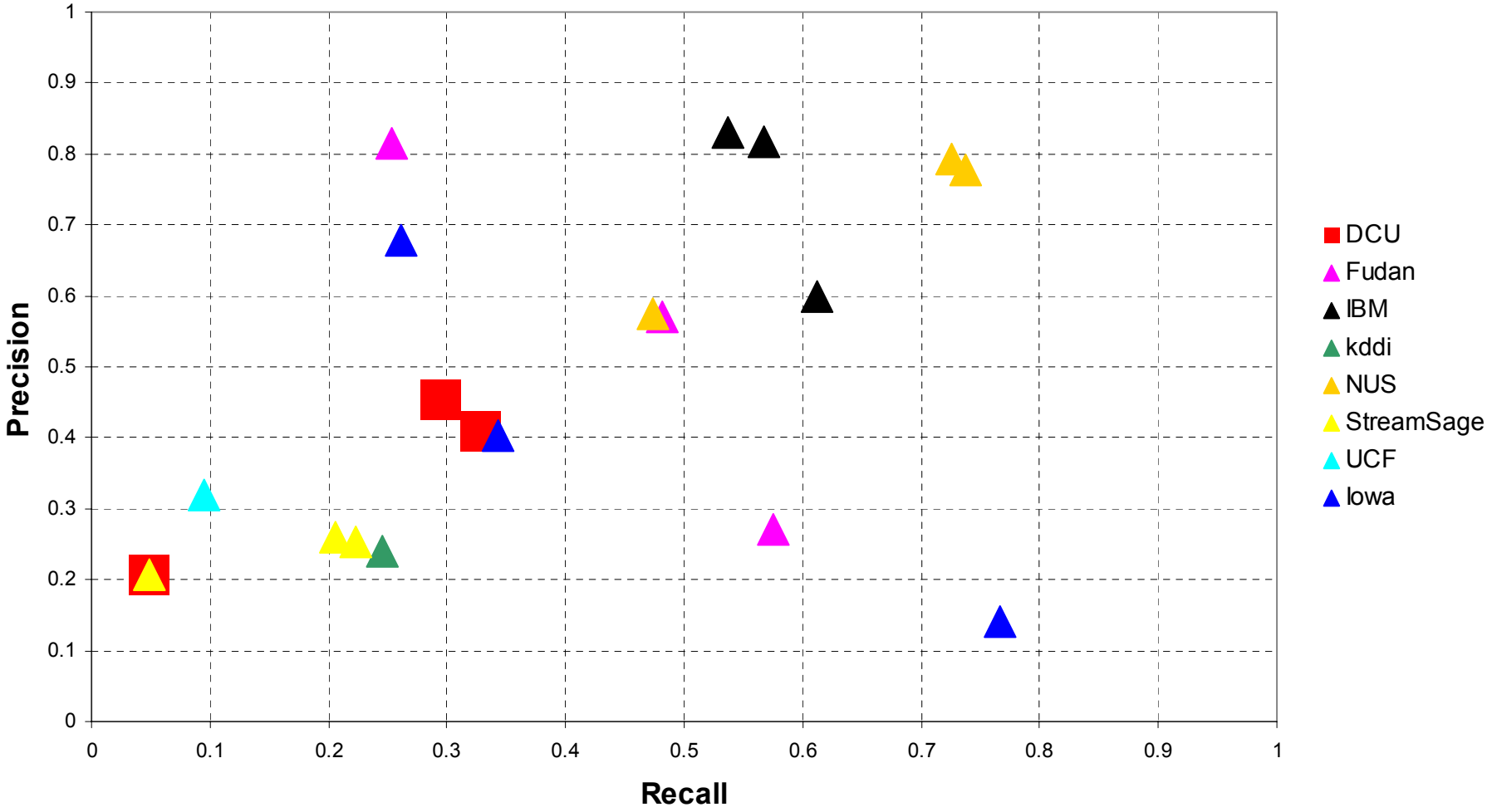
- 2 Additional Optional Runs
  - Specialised systems for ABC and CNN. Separate SVMs for each broadcaster **(DCU03_OPT_AV)**
  - Clustering algorithm in isolation **(DCU03_OPT_CLUSTER)**

DCU Results

CENTER FOR DIGITAL VIDEO PROCESSING

System ID	Recall	Precision
DCU03_REQ_AV	0.328	0.409
DCU03_REQ_AV_TEXT	0.294	0.453
DCU03_REQ_TEXT_ONLY	0.049	0.208
DCU03_OPT_AV	***0.313***	***0.453***
DCU03_OPT_CLUSTER	0.364	0.304

Overall Results - All Groups



Conclusions

CENTER FOR DIGITAL VIDEO PROCESSING

- 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


Thank You