A Pseudo-Statistical Approach to Commercial Boundary Detection

Prasanna V Rangarajan
Dept of Electrical Engineering
Columbia University

pvr2001@columbia.edu

1. Introduction

Searching and browsing through a news video archive in an effective manner presents a lot of challenges. A typical news program spans over a half hour, and there are no visible markers to help viewers find their way through the medium. It would be immensely useful to have an automated way of generating a list of topics addressed in the archived news material. An effective browsing interface for such a medium must also present the user with the option of skipping irrelevant content such as commercials and credits. Several approaches to news program analysis, indexing and retrieval can be found in recent literature. This report, in particular, addresses the issue of automatically detecting commercial boundaries from the bit stream of digitally captured news broadcasts.

The report is outlined as follows: Section 2 describes the structure of a commercial boundary; Section 3 describes the approach to detecting commercial boundaries and Section 4 provides samples results. Appendix A provides a list of programs developed for the purpose of commercial boundary detection. Appendix C provides snapshots & guidelines for using the tools developed for the above purpose.

2. Anatomy of a Commercial Boundary

Usually, commercial breaks are isolated from the actual news material by cues that television companies employ during the course of the broadcast. The most prominent ones include a series of black frames accompanied by a decrease in the audio volume prior to and subsequent to each commercial. Other cues include increased motion activity during the commercial, the absence of company logos, and the presence of anchorperson & background music. Some of these cues such as black and silent frames are easier to detect while the others are much harder to detect. The following figure illustrates a commercial boundary highlighting the presence of black frames coupled with periods of silence.

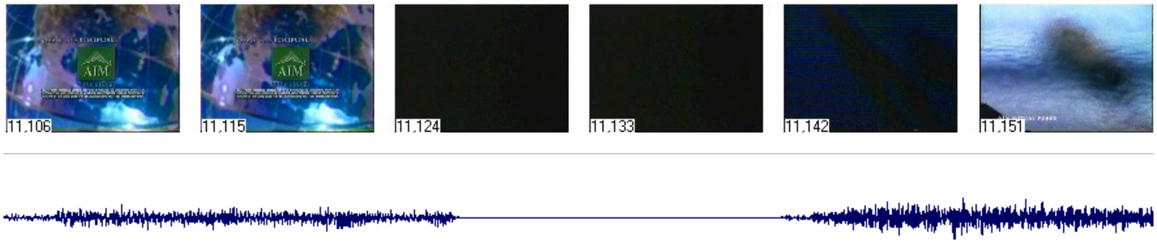


Fig 1.1 Commercial boundary illustrating the presence of black frames coupled with periods of silence

In contrast, it is more than likely that these prominent cues could occur during the course of the news broadcast, as shown in Fig 1.2. The interesting thing to note is that the black -frames are not associated with periods of silence. Further, the black frames are embossed with the logo of the broadcasting company, which might be difficult to spot in the following figure (Fig 1.2).

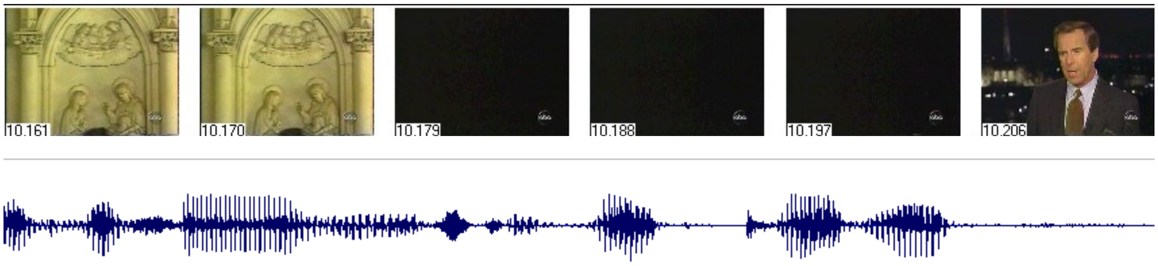


Fig 1.2 Non-Commercial boundary illustrating the presence of black frames & the absence of silent frames

3. Approach

A. FEATURE EXTRACTION

A survey of relevant literature revealed that attempts have been made in the past, to detect commercial boundaries based on the presence/absence of an anchorperson in the scene, the presence/absence of company logos and motion activity. However, these features did not prove to be very relevant for news videos in the TREC database. As for anchorpersons, sports updates in CNN news videos do not have a visible anchorperson in the scene. Motion activity proved to be an unreliable estimate due to the inferior quality of the motion estimation algorithm employed by the TREC encoder. The algorithm yields high motion estimates for consecutive black frames in a commercial boundary unlike the low motion activity observed in reality. The poor quality of the motion estimates can be attributed to the size of the search window employed during motion estimation. As for company logos, they do not appear until 5-10 seconds have elapsed since the resumption of the news broadcast. This time period is considerably large compared to the atomic nature of the boundary events.

For reasons mentioned above, black and silent frames proved to be the most reliable cues for detecting commercial boundaries in the TREC database. A preliminary study of the first order statistics of the luminance and the hue components revealed that black frames can be detected reliably by thresholding these features.

Feature Vector	Typical Value for a black frame
Mean Luminance	< 40.0
Mean Hue	< 40.0 degrees
Standard Deviation of Luminance	< 10.0

On the other hand, detecting if a video frame is embedded in silence is a far more involved task, partly because the sampling rates of the audio and the video data are different by several orders of magnitude. As before, the basic principle employed in detection is thresholding. The following figure (Fig 3.1) illustrates the manner in which audio samples are synced with the corresponding video frames, before silence detection is performed.

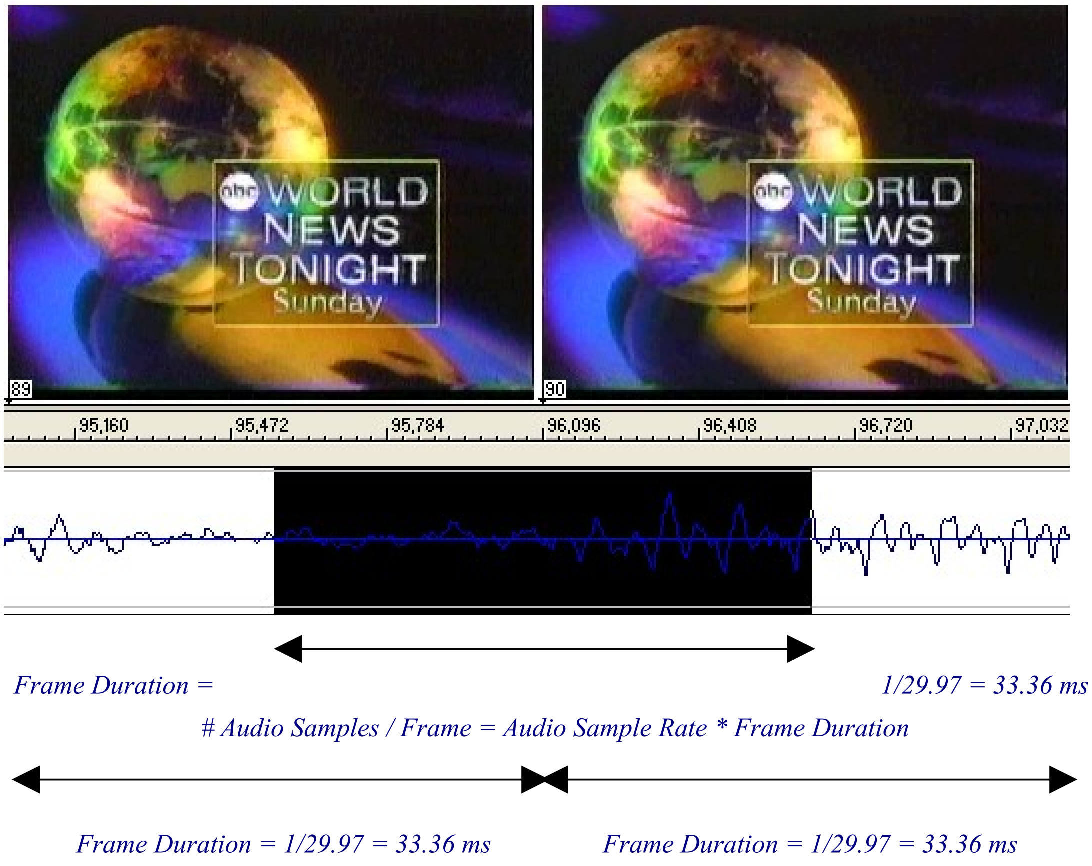

abc WORLD NEWS TONIGHT Sunday
89
90
95,160
95,472
95,784
96,096
96,408
96,720
97,032
Frame Duration =
1/29.97 = 33.36 ms
# Audio Samples / Frame = Audio Sample Rate * Frame Duration
Frame Duration = 1/29.97 = 33.36 ms
Frame Duration = 1/29.97 = 33.36 ms

Fig 3.1Syncning audio samples with their corresponding video frames

In order to detect if a video frame is embedded in silence, audio samples in the vicinity of the video frame [highlighted by black in the figure], are considered. At the video boundaries, only half the numbers of audio samples are considered for silence detection.

Feature Vector	Typical Value for a silent frame
Average Energy	< 130.0

Another audio feature that is widely used to reinforce the silence detection process is the average zero crossing rate (ZCR) in an audio frame.

The techniques presented above, require direct access to the audio and video data. This necessitates the use an MPEG decoder prior to feature extraction. DirectShow transform filters were designed to ensure that the process of feature extraction could be accomplished in real time. The algorithm can be altered suitably, to extract relevant features in the compressed domain.

B. PRE-PROCESSING

Commercial boundaries have a rather definite structure as outlined in Section 2. This makes it an ideal candidate for analysis using a generative hidden markov model. However, the atomic nature of these events lasting 10-30 frames, introduces reliability issues in the detector. The solution to this problem is partly motivated by approaches to similar problems in the speech recognition community. It involves consolidating feature information over a period of N frames, into units called *events*. In the present implementation, each event spans 20 frames with an overlap of 10 frames.

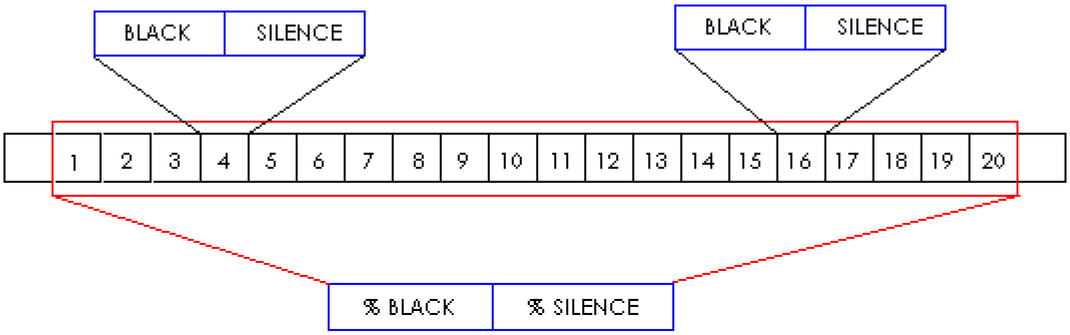
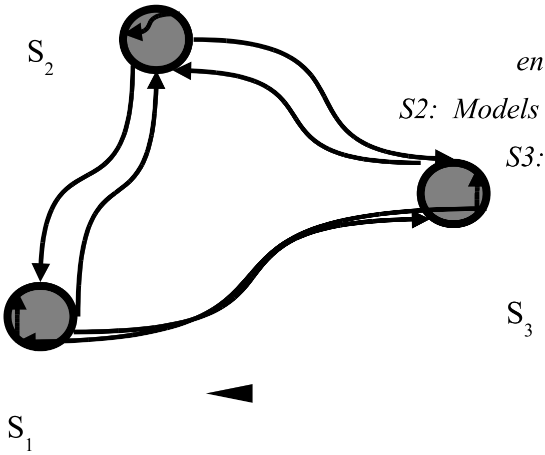


Fig 3.2 Pre-Processing of raw audio & video features

As illustrated in the above figure, the Boolean mid level features extracted at the frame level are consolidated in groups of 20 frames. The percentage of these Boolean mid level features is employed for the purpose of detection.

C. DETECTION

As outlined before, the transformation from a news story to commercial & vice-versa can be viewed as stochastic in nature and modeled using a hidden markov model. In the current implementation, a *3 state fully connected HMM with continuous observations* was used to realize the detector. The motivation for choosing 3 states is outlined below



S1: Models the entry into the transition

S2: Models the transition itself

S3: Models the exit from the transition

Fig 3.3 Generic 3-state fully connected HMM architecture

An increase in the number of states in the HMM does not yield a significant improvement in detection accuracy. The downside of choosing more states is that it requires more training examples to model a commercial boundary.

The feature vectors used in the training and evaluation phase included the %age of black frames in an event, %age of silent frames in the event and a quantized version of the time of occurrence of the event. The length of each training example was selected as 6 events, while the length of each test example was selected as 12 events. The lack of adequate training examples for modeling commercial boundaries necessitated the statistical modeling of all types of events except for commercial boundaries. As a result, commercial boundaries manifest themselves as strong local minima in the log likelihood curve as shown in the following figure.

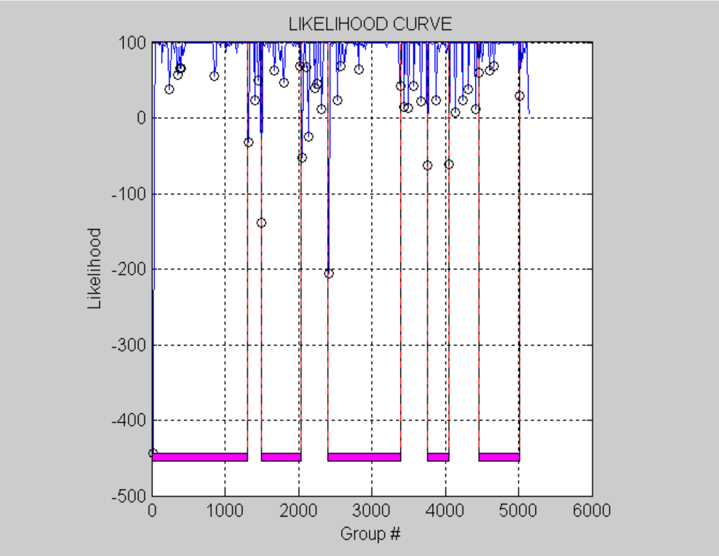


Fig 3.4 Log likelihood curve highlighting the news segments

The regions marked in pink represent the news segments while the unmarked regions represent commercial segments. On an average, there are 4-5 news segments per video, roughly separated by 2800 frames. These subtle details play a crucial role in eliminating several of the candidate commercial boundaries represented by an **o** in the figure shown above.

D. POST PROCESSING

Since the cues employed for boundary detection are also used to separate commercials themselves, post-processing of the local minima in the likelihood curve is crucial. Typically, the length of the cues for news ↔ commercial transition is longer than those for commercial ↔ commercial transitions. However, there were several instances of commercial boundaries, where the length of neighboring commercial ↔ commercial transition far exceeded those of the news ↔ commercial transition. This problem can be alleviated to a certain extent by choosing features such as motion activity and the presence/absence of a logo. In the present context, motion activity proved to be an unreliable estimate due to the inferior performance of the motion estimation algorithm employed by the encoder. As for using company logos, it does not appear until 5-10 seconds have elapsed since the resumption of the news broadcast. This time period is considerably large compared to the atomic nature of the boundary events. Hence an appropriate scaling algorithm needs to be employed.

The issues described above necessitated the use of a post-processing algorithm, which take into account broadcast production rules, to reliably detect the true commercial boundaries. The rules employed in the present algorithm, are summarized below.

Heuristic Production Rules	Typical Value (frames)
Minimum distance between news segments	3100
Maximum length of a news segment	10500

4. Results

Preliminary testing revealed that the features employed by the detector were salient enough to yield impressive recognition rates. The statistical model was trained on a single news video lasting 29.15 minutes, using the Baum Welch method. Exhaustive testing on a test set spanning 26 news videos and 236 commercial boundaries yielded the following precision and recall rates:

Description	Precision %	Recall %	# Missed	# False Alarms	# Boundaries
15 ABC News Videos	89.36	96.45	15	6	141
12 CNN News Videos	94.73	95.78	5	4	95
27 News Videos	91.52	95.76	20	10	236

The advantage of this approach over existing approaches to commercial detection is its underlying statistical nature. Most existing commercial detectors that boast higher recognition accuracies are either completely heuristic in nature and do not generalize well or have been evaluated in a limited framework with the purpose of classifying video segments into one of many classes.

An examination of the cases when the detector failed revealed rare scenarios where loss of transmission occurred resulting in black & silent frames or extremely short transition regions lasting a few frames or transition regions that are not primarily black.

5. References

R. Lienhart, C. Kuhknch & W. Effelsberg, "On the detection and recognition of television commercials", *Proc. of IEEE Int'l Conf. on Multimedia Computing and Systems, 1997*

S. Eickeler, S. Muller, "Context-Based Video Indexing of TV Broadcast News using Hidden Markov Models"

A. G. Hauptmann and Michael J. Witbrock, "Story Segmentation and Detection of Commercials in Broadcast News Video"

J. A. List, A. R. van Ballegooij, A. P. de Vries, "Known-Item Retrieval on Broadcast TV", *Report INS-R0104, 2001*

Z. Liu, Y. Wang and T. Chen, "Audio Feature Extraction & Analysis for Scene Segmentation & Classification"

J. Huang, Z. Liu and Y. Wang, "Joint Video Scene Segmentation and Classification based on Hidden Markov Model"

M. Slaney, D. Poncelon & J. Kaufman, "Mutlimedia Edges: Finding Hierarchy in all Dimensions", *Proc. of 9th ACM Int'l Conf. on Multimedia, 2001*

G. M. Snoek and Marcel Worring, "Multimodal Video Indexing: A Review of the State-of-the-Art", *2003*

L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proc. Of IEEE, 77(2), 1989*

Appendix A (List of programs & relevant information)

Program Name	Description
SetupCD.exe	Installation Program ftp://pvr2001:patti13@www.bannerless.com/SetupCD.exe
Batch_FX [Audio].exe	audio feature extraction in batch mode
Batch_FX [Video].exe	video feature extraction in batch mode
Comm_PreProcess.exe	pre-processing script
Comm_TrainHMM.exe	trains a continuous HMM to detect commercial boundaries
Comm_Detect.exe	detects commercial boundaries in a news video
Comm_BatchDetect.exe	batch mode commercial boundary detection program

Program Name	Input Format	Output Format
Batch_FX [Audio].exe	`*.WAV`	`*.AUD`
Batch_FX [Video].exe	`*.MPG`	`*.VID`
Comm_PreProcess.exe	`*.AUD + *.VID`	`*.CDF`
Comm_TrainHMM.exe	`*.CDF`	`-- NA --`

Comm_Detect.exe	`*.CDF + *.VID +`
	`*.CGT (Optional)`
		`*.CDO`

Comm_BatchDetect.exe	`*.CDF + *.VID`	`*.CDO`

File Format	Description
`*.WAV`	uncompressed audio stream extracted from MPEG news video
`*.MPG`	compressed news video (includes audio & video streams)
`*.AUD`	audio feature file
`*.VID`	video feature file
`*.CDF`	commercial detector feature file
`*.CDO`	commercial detector feature output

Note: The programs listed in the above table have an associated help button, which describes how to use the application.

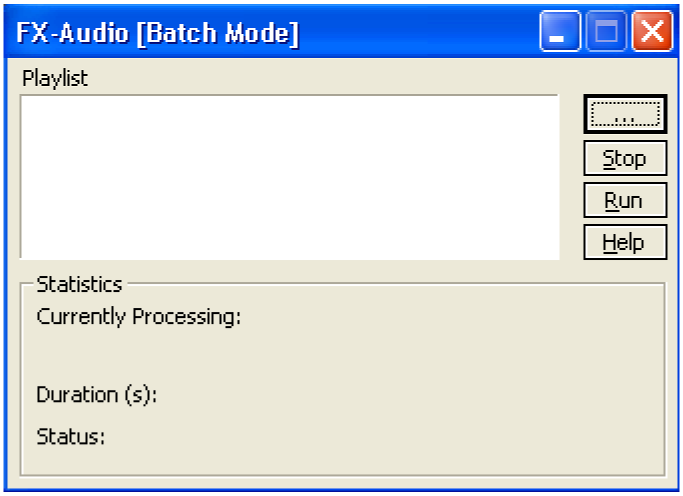
Appendix B (Supplementary Programs)

Some of the programs listed in Appendix A require additional programs to be installed before they can be successfully run. The following table lists those dependencies and links to download them where applicable.

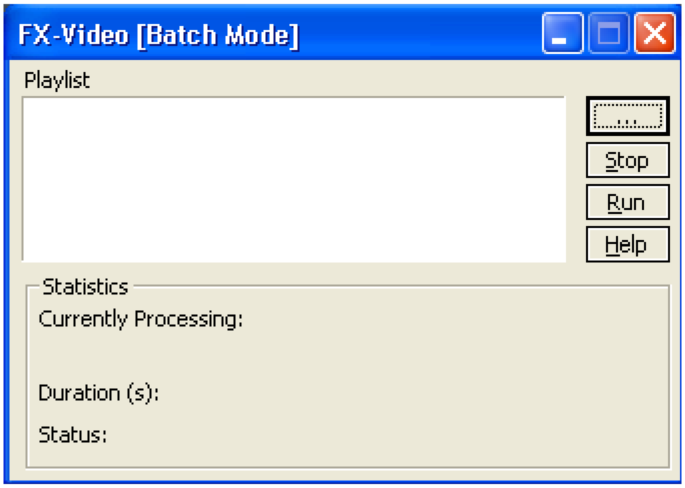
Program	Download Link (CTRL + click to follow link)
DirectX v9.0 Runtimes	http://download.microsoft.com/download/7/3/c/73cc71c0-13d9-4274-8d9c-33d8a528a396/dxwebsetup.exe
Intel IPP Runtimes	ftp://pvr2001:patti13@www.bannerless.com/Intel IPP.exe
Matlab v6.5 Runtimes	ftp://pvr2001:patti13@www.bannerless.com/mglinstaller.exe

Note: To download the above programs hold CTRL and click on the link

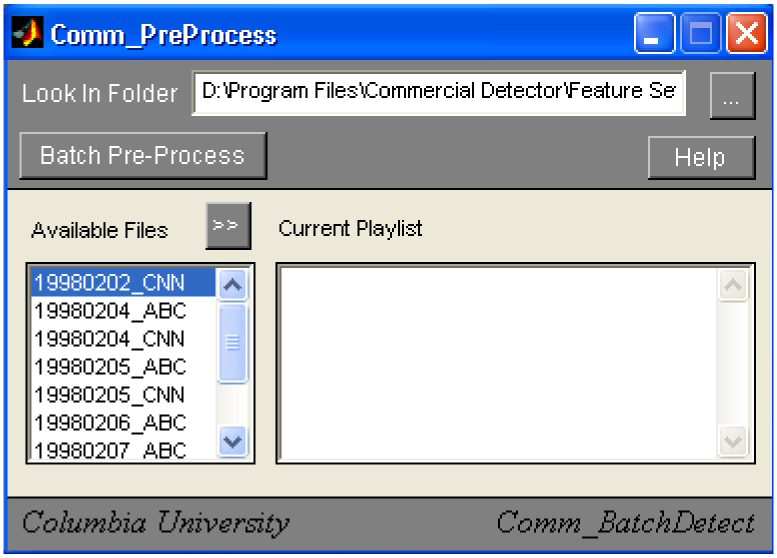
Appendix C (Screenshots)



- Click on the Ellipsis (…) button to add files to the current playlist
- Click on the Run button to start feature extraction
- Click on the Stop button to stop feature extraction



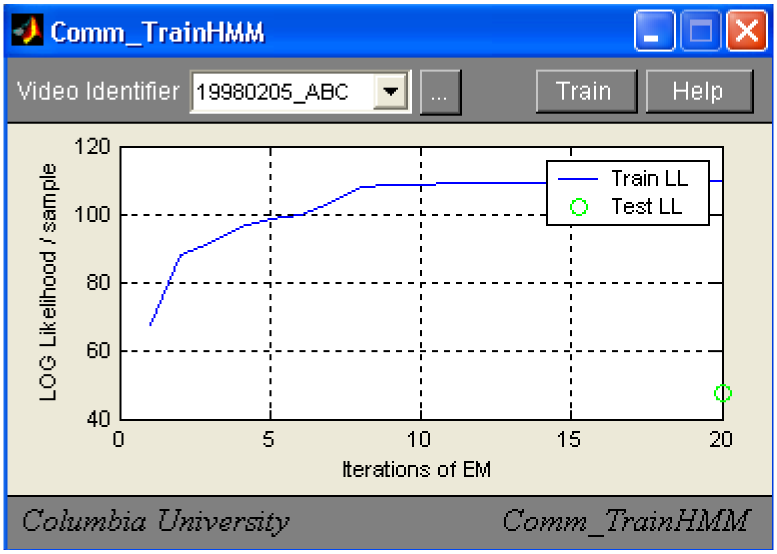
- Click on the Ellipsis (…) button to add files to the current playlist
- Click on the Run button to start feature extraction
- Click on the Stop button to stop feature extraction



Look In Folder: D:\Program Files\Commercial Detector\Feature Se

19980202_CNN
19980204_ABC
19980204_CNN
19980205_ABC
19980205_CNN
19980206_ABC
19980207_ABC

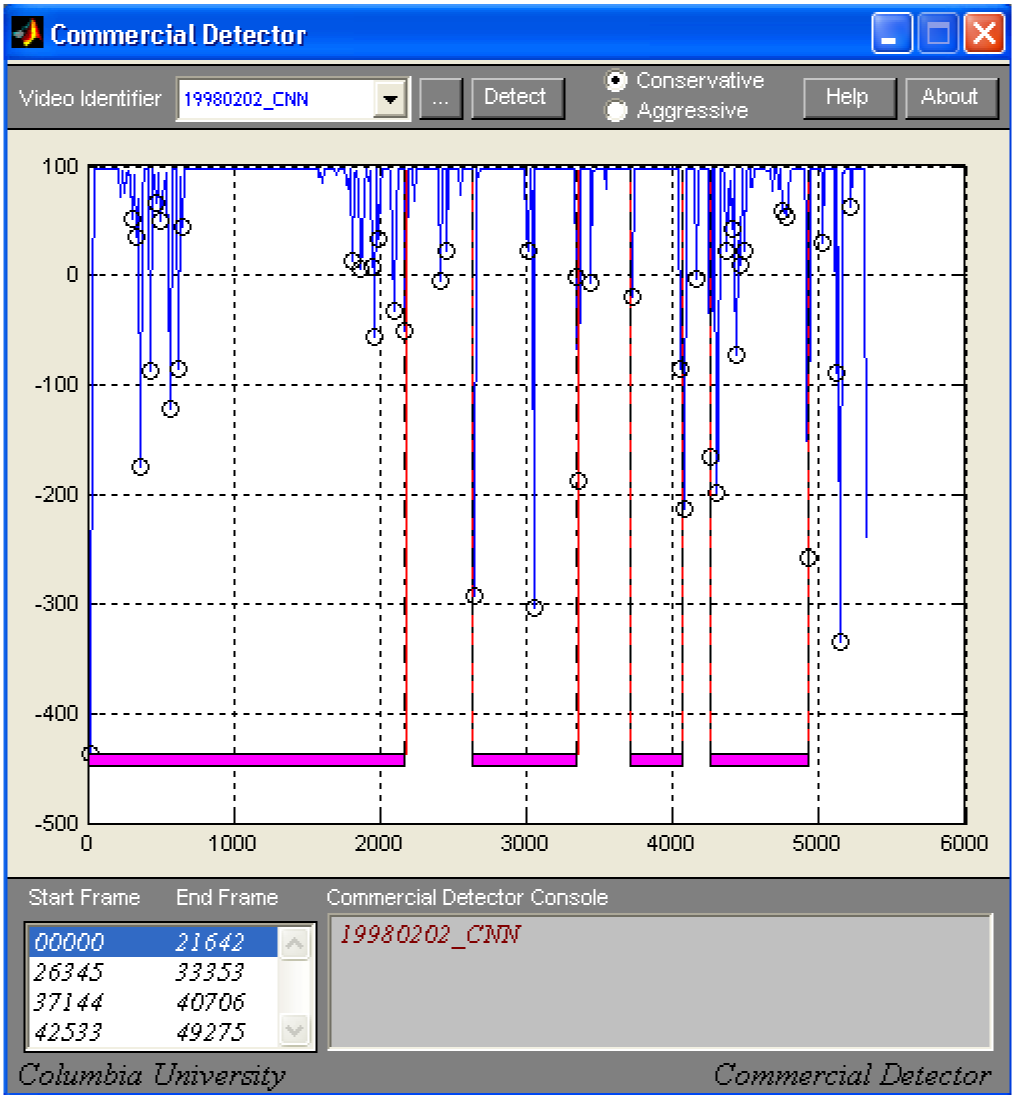
- Click on the Ellipsis (…) button to populate the list control with feature files (*.AUD + *.VID) in the selected folder
- Now, select files from the list control and click the (>>) button to add these files to the current playlist
- Click on the Batch Pre-Process button to pre-process the selected feature files (*.AUD + *.VID)



Video Identifier 19980205_ABC

- Click on the Ellipsis (…) button to populate the Video Identifier control with pre-processed feature files (*.CDF) in the selected folder

- Click on the Train button to train a Mixture of Gaussians Hidden Markov Model to detect all events except commercial boundaries



Video Identifier 19980202_CNN

● Conservative
○ Aggressive

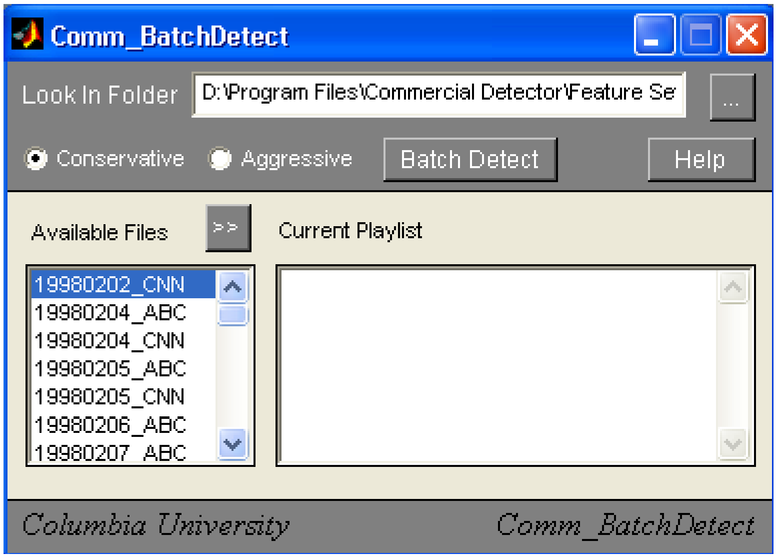
Start Frame	End Frame
00000	21642
26345	33353
37144	40706
42533	49275

19980202_CNN

- Click on the Ellipsis (…) button to populate the Video Identifier list control with .CDF files in the selected folder

- Select a file from the Video Identifier list control and click the Detect button to detect commercial boundaries in the selected CDF file

Note: By default, the detector runs in Conservative mode which supports higher precision and lower recall rates, while the Aggressive mode supports lower precision and higher recall rates.



Look In Folder: D:\Program Files\Commercial Detector\Feature Se

● Conservative ○ Aggressive

19980202_CNN
19980204_ABC
19980204_CNN
19980205_ABC
19980205_CNN
19980206_ABC
19980207_ABC

- Click on the Ellipsis (…) button to populate the list control with CDF files in the selected folder
- Now, select files from the list control and click the (>>) button to add these files to the current playlist
- Click on the Batch Detect button to detect commercial boundaries for files in the current playlist

Note: By default, the detector runs in Conservative mode which supports higher precision and lower recall rates, while the Aggressive mode supports lower precision and higher recall rates.