Audio and Video Processing for Automatic TV Advertisement Detection

Seán Marlow, David A. Sadlier, Karen McGeough, Noel O'Connor, Noel Murphy.
Visual Media Processing Group,
Dublin City University,
Ireland.

email: *marlows@eeng.dcu.ie* ***phone:*** *+353-1- 7005120*

Abstract

As a partner in the Centre for Digital Video Processing, the Visual Media Processing Group at Dublin City University conducts research and development in the area of digital video management. The current stage of development is demonstrated on our Web-based digital video system called *Físchlár* [1,2], which provides for efficient recording, analyzing, browsing and viewing of digitally captured television programmes. In order to make the browsing of programme material more efficient, users have requested the option of automatically deleting advertisement breaks.

Our initial work on this task focused on locating ad-breaks by detecting patterns of silent black frames which separate individual advertisements and/or complete ad-breaks in most commercial TV stations. However, not all TV stations use silent, black frames to flag ad-breaks. We therefore decided to attempt to detect advertisements using the rate of shot cuts in the digitised TV signal. This paper describes the implementation and performance of both methods of ad-break detection.

1. INTRODUCTION

1.1 The Físchlár System

The imminent rapid expansion in the number of TV channels is driving the need for efficient digital video indexing, browsing and playback systems. For the past three years, the Centre for Digital Video in DCU has been working towards the provision of such a system. The current stage of development is demonstrated on our Web-based digital video system called *Físchlár* [1,2], which is now in hourly use by over 1000 registered users. At present a user can pre-set the recording of TV broadcast programmes and can choose from a set of different browser interfaces which allow navigation through the recorded programmes. As our research develops we will plug in increased options such as personalisation and programme recommendation, automatic recording, SMS/WAP/PDA alerting, searching, summarising and so on.

To initiate the recording of a programme, a user browses the TV schedule and selects those programmes to be recorded - our system will then automatically record (digitally) that programme at broadcast time, much the same as a home VCR. After we record a programme, we then automatically segment it using our shot boundary detection technique based on colour histogram comparison, so that the content becomes easily browsable through our various user interfaces. The analysed programme is then added to our archive of recorded programmes which a user can scroll through and then select one for browse/playback. As a user browses through a programme he/she can then stream the video to their desktop.

In order to more efficiently browse and view programme content, many users have requested the option of skipping the ad-breaks, which have accompanied commercial TV programmes since the mid 1940s. To do this, the general characteristics of advertisements must be examined.

1.2 Characteristics of Advertisements

On most commercial TV stations, ad-breaks are flagged by a series (of varying length) of black, silent frames at the beginning and end of each ad-break. Individual advertisements are usually separated by a shorter sequence of black, silent frames. Detection of ad-breaks using this pattern of separators is described in Section 2 and results for this approach are presented in Section 3.

However, some TV stations (TV3 and Channel 4 in Físchlár) do not use black, silent frames to flag ad-breaks. Therefore some additional characteristics of advertisements need to be utilised. In order to maximise the visual impact of TV advertisements, producers frequently use a faster rate of shot cuts than normal TV content [3]. The use of shot cut rate for locating ad-breaks is detailed in Section 4 and results are given in Section 5.

2. AD-BREAK DETECTION USING TRANSMITTED FLAGS

2.1 MPEG Bitstream Processing

The *Físchlár* system captures television broadcasts and encodes the programmes according to the MPEG-1 digital video standard with the audio signal coded in line with the Layer-II profile [4]. An inherently dark or 'black' frame of a video may be recognised via examination of the frame's luminance histogram, where most of the 'power' is at the bottom end of the pixel amplitude spectrum. Thus, by comparing an average pixel value, representing an entire frame, against some given threshold, a decision on whether that frame may be considered 'black' or not, may be made. Furthermore, a depression in audio volume for a particular video frame may be recognised as follows: a summation of the absolute value of all the audio samples corresponding to one video frame may be defined as the 'audio level' for that frame, i.e. for a relatively silent frame a low audio level would be expected. Thus, by comparing this audio level to some threshold, silent frames may be detected.

The above-mentioned method is a straight-forward approach to the task of locating, within a television programme, groups of black, silent frames, which provides for automatic detection of advertisement breaks. However, the method does require direct access to both video pixels and audio samples. Therefore it necessitates a full decode of the captured programme from its compressed format, which is highly undesirable from a computational point of view. It was thought that the same assessment and classification of the individual frames of a captured television signal might be more efficiently made as follows:

For video: an examination of the DC Discrete Cosine Transform coefficients of a frame, which represent the weight of its zero-frequency content, with a view to establishing whether or not the frame is inherently dark enough to be labeled 'black'.

For audio: an inspection of the weight of the scalefactors [4] of the signal's (low) frequency subbands with a view to establishing whether or not a video frame's accompanying audio signal power is low enough for it to be labeled 'silent'.

2.2 Black Frame Detection

An MPEG-1 video frame is divided into slices, which are subdivided into macroblocks which each contain 6 blocks (four luminance, two chrominance) of (8x8) pixels transformed by a 2D-DCT [4]. The four luminance blocks ("Y-blocks") provide the essential information on how dark or 'black' a frame effectively is. Each Y-block consists of a DC coefficient, which represents its mean luminance intensity, and a number of AC coefficients, which represent its non-zero frequency content. The DC value corresponds to an average intensity value for each block. It was thus assumed that a decision on the inherent darkness of a block could be made with acceptable accuracy, via examination of the DC coefficients exclusively. The average luminance intensity value for each frame was determined from the DC-DCT coefficients provided within individual Y-blocks. This value was expected to be relatively low for inherently dark frames and higher for brighter frames. The mean DC-DCT value for the whole clip was calculated by averaging over all individual frames. The 'black-frame' threshold was then found by trial and error examination of various ad-break clips. The figure which gave the most consistent results was: $\mathbf{thresh_v = 0.48 * DC\text{-}DCT_{avg}}$

2.3 Silent Frame Detection

Físchlár encodes television audio signals according to the MPEG-1 Layer-II compression algorithm. Scalefactors are used to scale each group of 12 samples in each subband such that they use the full range of the quantiser. The scalefactor for such a group is determined by the next largest value (given in a look-up table) to the maximum of the absolute values of the 12 samples. Thus the scalefactor provides an indication of the maximum power exhibited by any one of the 12 samples within the group. The audio power level for each video frame was determined by superposition of the scalefactors corresponding to the groups of audio samples to which the frame is associated. By thresholding this value, a decision on whether the frame is 'silent' or not may be made. The overall mean audio level value was calculated by averaging over all individual frames. The 'silent-frame' threshold was then determined by trial and error examination of various ad-break clips and the figure which gave the most reliable results was: $\mathbf{thresh_a = 0.073 * audio\_level_{avg}}$

2.4 Recognition of Advertisement Breaks

As explained, the occurrence of simultaneous black-frames/audio-depressions may indicate the existence of an ad-break. However, it is possible that these indicators also occur during the programme itself. For example, they are not uncommon when News programmes cut back and forth from anchorperson to news reports, or during scene changes during a soap opera. To combat this problem, and its consequence of detection/removal of valuable programme content, some strict conditions had to be enforced.

It was noted that the flags occurring between individual advertisements tended to be of at least 6 frames in length. Thus to aid against detection of black-frame/audio-depression occurrences not associated to ad-breaks which may sporadically occur during programmes, it was decided to recognise them only if they exhibit a series of **at least 6 consecutive 'black/quiet' frames**.

Upon examination of 20+ advertisement breaks from various television stations, the longest advertisement recorded lasted 76secs, with approximate average advertisement duration of 25secs. Thus it was decided that upon detection of a series of (at least 6) 'black/quiet' frames, if another distinct series was not detected within a **window of 90 seconds**, then the initial series must therefore not correspond to an ad-break and should be ignored.

Examination of 20+ advertisement breaks from various television stations revealed that the minimum number of individual advertisements within an ad-break was four. It was decided, to further prevent against the possibility of relevant programme material being mistakenly recognised as advertisement, that the **recognition process would not succeed if the number of advertisements within one ad-break was less than three**.

3 RESULTS FOR AD-BREAK DETECTION USING TRANSMITTED FLAGS

3.1 Test material

Físchlár provided 10 short television programme clips in MPEG-1 format. The recordings were chosen such that they consisted of programme material with at least one complete ad-break somewhere in the middle. The procedures in Section 2 were executed on all 10 clips. The results achieved, together with the manually determined record of the true location of the ad-breaks within each clip, are shown in Table 1.

I.D./TV Channel of clip	Length of clip (frames/mins)	Ad-break detected (frame - frame)	True ad-break location (frame-frame)
(a) UTV	18009 frames (˜ 10mins)	5795 - 8435	5794 - 8430
(b) RTE1	30002 frames (˜ 20mins)	2829 - 8777	2832 – 8771
(c) RTE1	30003 frames (˜ 20mins)	19475 - 23368	19477 – 23366
(d) TG4	48012 frames (˜ 30mins)	4923 - 7261	4923 – 7262
(e) TG4	40512 frames (˜ 25mins)	17804 - 20842	17803 - 20846
(f) Net2	48018 frames (˜ 30mins)	21287 – 25996 and 37745 - 44278	21289 – 25999 and 37748 - 44280
(g) RTE1	30001 frames (˜ 20mins)	15065 - 19249	15065 - 19760
(h) Net2	43509 frames (˜ 30mins)	18706 - 22066	18706 - 22574
(i) UTV	33009 frames (˜ 20mins)	12878 - 14203	12880 - 14808
(j) UTV	18009 frames (˜ 10mins)	7737 - 10408	7704 - 10411

Table-1 Location of true and detected ad-breaks

[**Note:** Clip **(f) Net2** has two ad-breaks. Thus we are dealing with 10 clips incorporating 11 ad-breaks.]

These detailed results are summarised in graphical form in Figure 1.

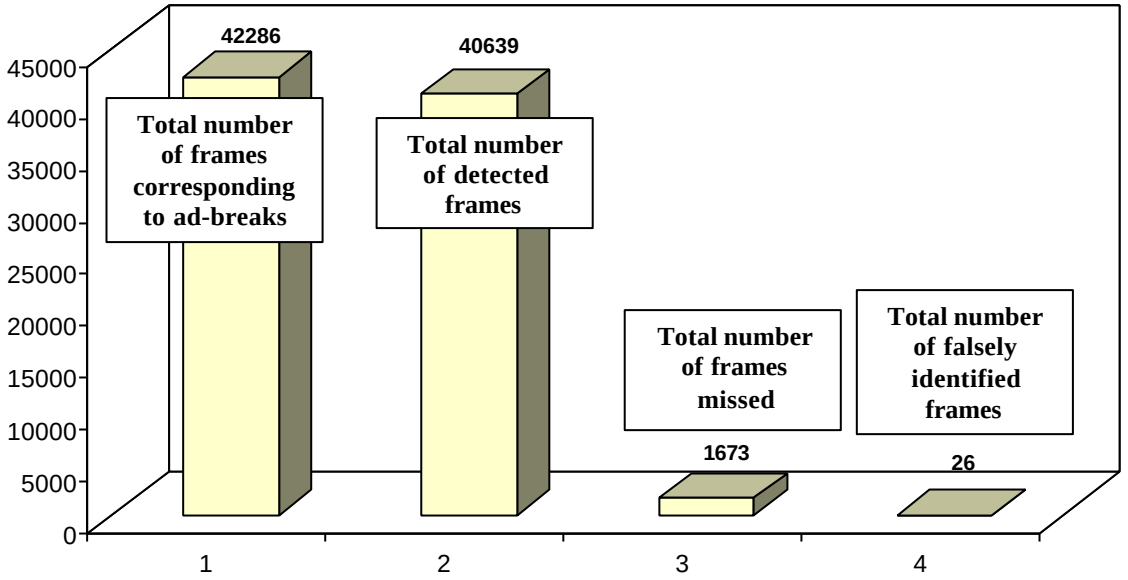

Figure-1 Total number of frames corresponding to ad-breaks, detected ad-breaks, missed frames and falsely identified frames.

3.2 Precision & Recall

To evaluate the performance of the ad-break detector, we employ two standard information retrieval metrics:

Precision is a percentage showing how accurate the system is at detecting frames that correspond only to ad-breaks.

$$\text{Precision} = \frac{100 * [\text{Actual no. frames in ad-break} - \text{No. frames missed}]}{[\text{Actual no. frames in ad-break} - \text{No. frames missed} + \text{No. falsely identified frames}]}$$

Recall measures the percentage of frames corresponding to actual ad-breaks that the system has detected.

$$\text{Recall} = \frac{100 * [\text{Actual no. frames in ad-break} - \text{No. frames missed}]}{[\text{Actual no. frames in ad-break}]}$$

These performance metrics for all 10 test clips are presented in Table-2.

Clip	Precision	Recall
(a) UTV	99.8	99.9
(b) RTE1	99.8	100
(c) RTE1	99.9	100
(d) TG4	100	99.9
(e) TG4	99.9	99.8
(f) Net2 (1)	99.9	99.9
(f) Net2 (2)	99.9	99.9
(g) RTE1	100	89.1
(h) Net2	100	86.8
(i) UTV	99.8	68.6
(j) UTV	100	98.7

Table-2 Precision and Recall figures for test material

4. AD-BREAK DETECTION USING SHOT CUT RATE

The results in Section 3 are very promising, for those TV channels which use silent, black frames to delineate advertisements. Unfortunately, in Físchlár, TV3 and Channel 4 do not provide these flags. Therefore, alternative ad-break detection methods are currently being investigated.

4.1 Shot Cut Rate

We noticed that the rate of shot cuts is often increased to maximise the visual impact of ads. This observation has been also noted by researchers on the Informedia Digital Library project who reported that the rate of shot cuts for an advertisement rises above 1.7 times the mean rate of shot cuts for the whole program [3].

The average rate of shot-cuts over three entire MPEG video files containing ad-breaks was calculated and the results are tabulated in Table 3.

Name of MPEG file	Shot Length	Average Rate of Shot-cuts(Hz)
Good Morning Vietnam	90.09 frames	0.27777
Free Willy	67.08 frames	0.37268
Thelma and Louise	97.71frames	0.2558

Table 3 Average Rate of Shot-cuts in Test Sequences

The average shot length over the 3 sequences is 85 frames. In our initial tests the shot length threshold for ad-break detection was set at 85/1.7 = 50 frames. The actual shot length is smoothed by averaging over 20 shots and the minimum number of consecutive shots whose length must be less than 50 frames is set at 10.

5. RESULTS FOR AD-BREAK DETECTION USING SHOT CUT RATE

5.1 Initial Tests

Físchlár supplied 3 test sequences which had shot cuts marked. The location (in Frame Nos) of the actual and detected ad-breaks are listed in Table 4 where X in a Detected row indicates ad-breaks which were missed and X in an Actual row indicates ad-breaks which were falsely detected.

Name of Test Sequence		1st Ad Break		2nd Ad Break		3rd Ad Break	
		Start	End	Start	End	Start	End
Thelma & Louise	Actual	39150	46250	90750	99050	145825	153600
	Detected	39424	45074	91585	97544	145999	152329
G'Morning Vietnam	Actual	50200	56550	94825	100825	145525	151825
	Detected	X	X	X	X	140614	148592
Free Willy	Actual	63875	70100	100625	108425	X	X
	Detected	64192	69941	101729	107698	133209	138669

Table 4 Location of Actual and Detected Ad Breaks using Shot Length

The Precision and Recall values for the three test sequences, as defined in Section 3.2, are listed in Table 5

Film	Precision %	Recall %
Free Willy	100	83.0
Good Morning Vietnam	34	31
Thelma & Louise	92.62	85

Table 5 Precision and Recall Values for 3 Test Sequences using Shot Length

6. CONCLUSIONS & CONTINUING WORK

Precision figures of over 99.7% for the first detection method over all clips indicates excellent performance in the prevention of false positives, which is highly desirable, as the user won't tolerate loss of wanted material. In all except 3 clips the Recall figure was over 99.6% indicating that very few ads were missed. In the other 3 the final black/silent flag lasted less than our threshold of 6 consecutive frames. Thus the final ad in the ad-break was not detected.

Our initial results for the location of ad-breaks using shot length indicate that this method is much less reliable. Work is continuing to increase accuracy by optimising thresholds but this is expected to produce only marginal improvement. Table 3 shows a wide variability in the average shot cut rate for different programs and indicates the need for an adaptive threshold to locate ad-breaks. Also, work on a related project on locating highlights in movies, indicates that there is a greater level of "visual activity", as measured by accumulated differences between consecutive frames, in shots within ad-breaks.

REFERENCES

1. *Centre For Digital Video Processing/Físchlár Website*. http://lorca.compapp.dcu.ie/Video/

2. Lee H, Smeaton A, O'Toole C, Murphy N, Marlow S and O'Connor N. *The Físchlár Digital Video Recording, Analysis, and Browsing System*, RIAO 2000 - Content-based Multimedia Information Access. Paris, France, 12-14 April 2000.

3. A.G. Hauptmann and M.J. Witbrock. Story Segmentation and Detection of Commercials in Broadcast News Video. Proceedings of Advances in Digital Libraries Conference, Santa Barbara, CA., April 22-24, 1998, 168-179.

4. Rao, K.R. and Hwang, J.J. *Techniques & Standards for Image, Video and Audio Coding*. Prentice Hall, 1996