AUDIO/VISUAL ANALYSIS

FOR

HIGH-SPEED TV ADVERTISEMENT DETECTION FROM MPEG BITSTREAM

David A. Sadlier B. E. (Electronic)
January 2002

MASTER OF ENGINEERING

IN

DIGITAL AUDIO/VIDEO PROCESSING

Supervised by Dr Noel O'Connor

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr Noel Connor, Dr Sean Marlow & Dr Noel Murphy for their extensive guidance and commitment to this project. Thanks also to all other friends/colleagues for their contribution to the establishment of such an inspiring and communal work environment that is the Visual Media Processing Group.

DECLARATION

I hereby declare that, except where otherwise indicated, this document is entirely my own work and has not been submitted in whole or in part to any other university.

Signed: ..... Date: 2002

ABSTRACT

Advertisement breaks during or between television programmes are typically flagged by series of black-and-silent video frames, which recurrently occur in order to audio-visually separate individual advertisement spots from one another It is the regular prevalence of these flags that enables automatic differentiation between what is programme content and what is advertisement break. Detection of these audio-visual depressions within broadcast television content provides a basis on which advertisement detection may be achieved This document reports on the progress made in the development of this idea into an advertisement detector system that automatically detects the advertisement breaks directly from the MPEG-1 encoded bitstream of digitally captured television broadcasts

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF GRAPHS

TABLE OF TABLES

TABLE OF ABBREVIATIONS

AC	'Alternating Current' – refers to non-zero-frequency
B-frame	Bidirectionally predicted frame
DC	'Direct Current' – refers to zero-frequency
DCT	Discrete Cosine Transform
Fps	Frames per Second
GOP	Group of Pictures
I-frame	Intra-coded video frame
ISO	International Standards Organisation
MAV	Mean Audio Volume
MLI	Mean Luminance Intensity
MPEG	Moving Pictures Experts Group
PDA	Personal Digital Assistant
P-frame	Predicted video frame
SMS	Short Messaging Service
VCR	Video Cassette Recorder
WAP	Wireless Application Protocol
Y-block	Luminance Blocks

1. INTRODUCTION

1.1 BACKGROUND

1.1.1 Visual Media Processing Group (VMPG)

Originally known as the *Video Coding Group*, the *Visual Media Processing Group (VMPG)* of the School of Electronic Engineering was founded in 1991 by two senior academic staff in order to carry out high quality basic research in the area of video compression In recent times, the group's focus and associated expertise has expanded to incorporate many other areas of image, video and recently audio processing Currently, the group consists of three academic staff members of the School of Electronic Engineering who co-ordinate the work of the group, seven full-time post-graduate students carrying out basic research, and a steady turn-over of undergraduates carrying out short-term student projects Some of the projects under current study relate to multimedia applications and come under the headings of sports summarisation, news story segmentation, high-speed shot boundary detection, object segmentation, speech/speaker discrimination

1.1.2 Centre for Digital Video Processing: *Físchlár*

The *Visual Media Processing Group* is a co-constituent of *The Centre for Digital Processing The Centre* is a cross-disciplinary research foundation which conducts concentrated research in developing innovative technologies fundamental to the realisation of efficient video content management The current stage of development is demonstrated in the web-based digital video system, *Físchlár* [1]

Físchlár provides for efficient recording, analysing, browsing and viewing of digitally captured television programmes At present a user can pre-set the recording of TV

[1] Lee H, Smeaton A, O'Toole C, Murphy N, Marlow S & O'Connor N, *The Fischlar Digital Video Recording, Analysis and Browsing System*, Proc Content based Multimedia Information Access (RIAO 2000), Vol 2, pp 1390-1399, Paris, France, 12 14 April 2000

broadcast programmes and can choose from a set of different browser interfaces which allow navigation through the recorded programmes As the research develops, increased options such as personalisation and programme recommendation, automatic recording, SMS/WAP/PDA alerting, searching, summarising etc are being plugged in The architecture of the *Físchlár* system is illustrated in **Figure 1-1**

1.1.3 Using *Físchlár*

To initiate the recording of a programme, a user browses the TV schedule and selects those programmes to be recorded - *Físchlár* will then automatically record and encode, according to the MPEG-1 digital video standard (see **Chapter-3**), that programme at broadcast time, much the same as a home VCR.
After a programme is recorded, it is then automatically segmented using a shot boundary detection technique based on colour histogram comparison, so that the content becomes easily browsable through the various user interfaces The analysed programme is then added to the archive of recorded programmes through which a user can scroll and then select one for browse/playback. As a user browses through a programme he/she can then stream the video to their desktop See **Figure 1-2**

1.2 Introduction To Advertisement Detection

1.2.1 Motivation

For the development of a more efficient video browsing/viewing tool for digitised television, it is desirable to present the user with the option of skipping irrelevant content A television programme is typically accompanied by beginning/end credits with one or more ad-breaks somewhere in the middle To the user, these features of a programme would be generally regarded as an insignificant part of the material

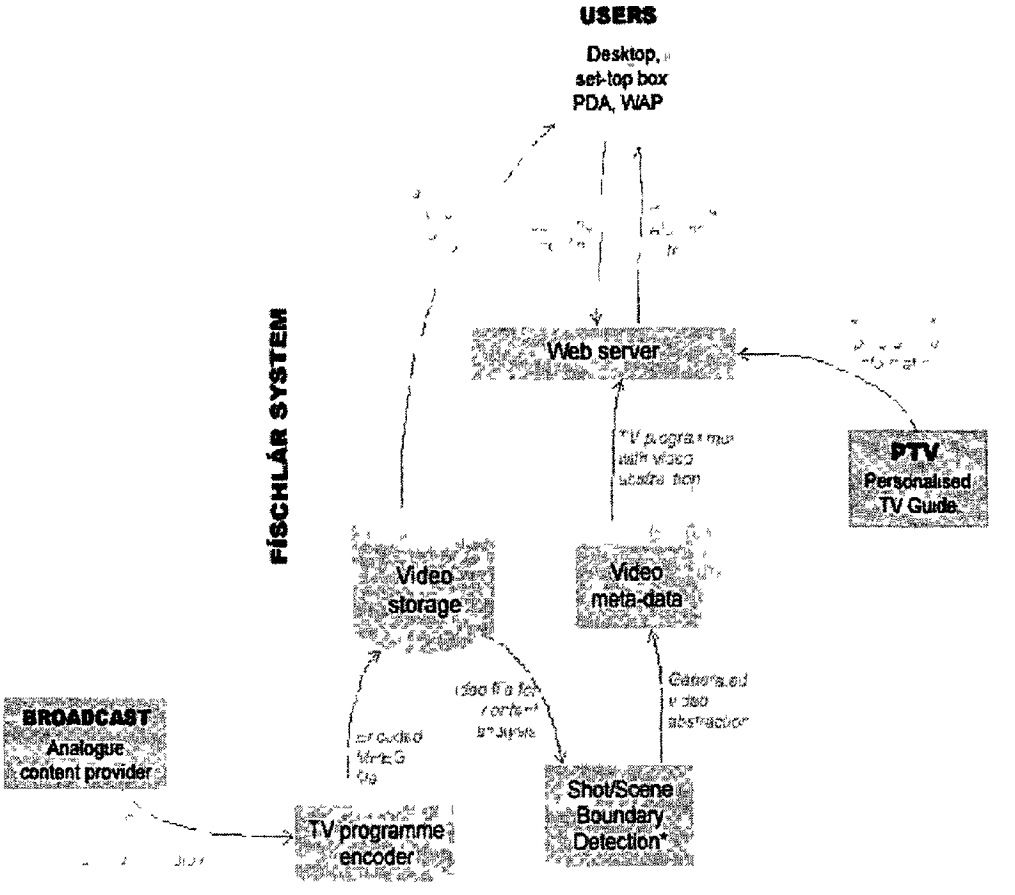


Figure 1-1 *Fischlar* architecture

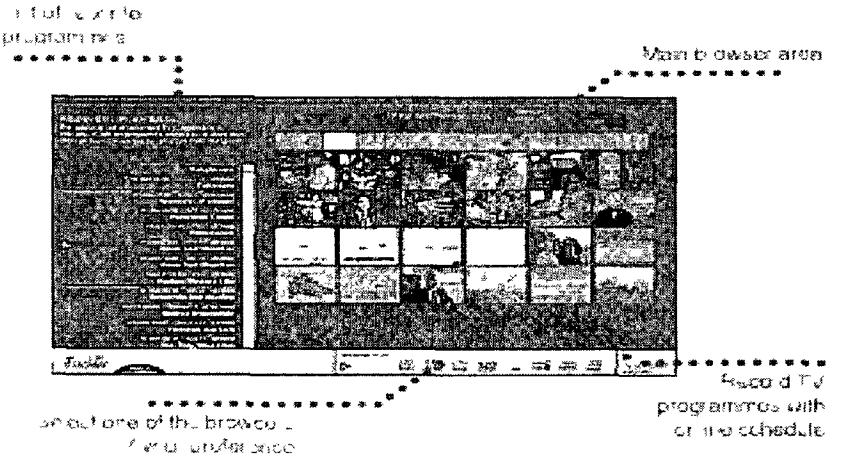


Figure 1-2 *Fischlar* screen demonstration

Hence, if automatically locating the ad-breaks was possible, the efficiency of programme browsing/viewing could be increased

1.2.2 Advertisement Detection

Upon manual investigation of the broadcasting traits of the available Irish terrestrial television stations which transmit advertisment breaks, it was noted that in most cases, the individual advertisement spots within any break are delimited by series of very dark (black) video frames simultaneously accompanied by a depression in the audio signal (silence) These features are characteristic of the typical ad-break production specification which argues that it is desirable to have each individual advertisement spot audio-visually distinguishable from its neighbouring spots See **Figure 1-3**

Detection of these features, coupled with some pattern recognition relating to typical ad-break structure, would provide a basis on which the location of advertisement breaks within entire broadcast television programmes could be precisely detected

1 2 2 1 Definition Black Video Frame

From this point forward it is important to be clear what is meant by the 'black frame' label The definition of a **black video frame** is one which, as a whole, is intrinsically dark with a uniform pixel value throughout and contains no meaningful content

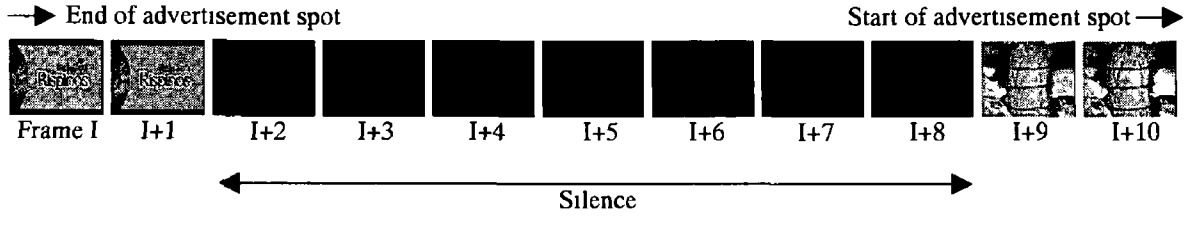


Figure 1-3 Two individual advertisement spots separated by a series of **black** video frames simultaneously accompanied by **silence**

Black frames in television programmes typically manifest themselves as visual separators between advertisement spots, which make the shot transitions less visually disturbing However, they may also sporadically occur within the programme or ad-break content For example, they are sometimes found providing a pausing link between two successive scenes, or simply within the content of a nightfall scene

1 2 2 2 Definition Silent Video Frames

A **silent** video frame is one for which the absence of significant audio information is aurally noticeable (apart from low volume noise) Silent video frames occur sporadically throughout television content

1 2 2 3 Black Video Frame Detection

A **black** video frame may be recognised by its luminance histogram, which would be typically characterised by having most of its 'power' at the bottom end of the pixel amplitude spectrum, corresponding to black or very dark pixels

For example, **Figure 1-4** demonstrates how trivial a task it is to discriminate an inherently dark frame from a more generic frame when the video pixel information is available

1 2 2 4 Silent Video Frame Detection

A summation of the absolute value of all the individual audio samples corresponding to the temporal length of one video frame may be defined as the 'audio level' for that frame, i e for a video frame with relatively quiet audio, a low audio level would be expected Thus, by thresholding this audio level, **silent** video frames (of intensity defined by threshold) may be detected

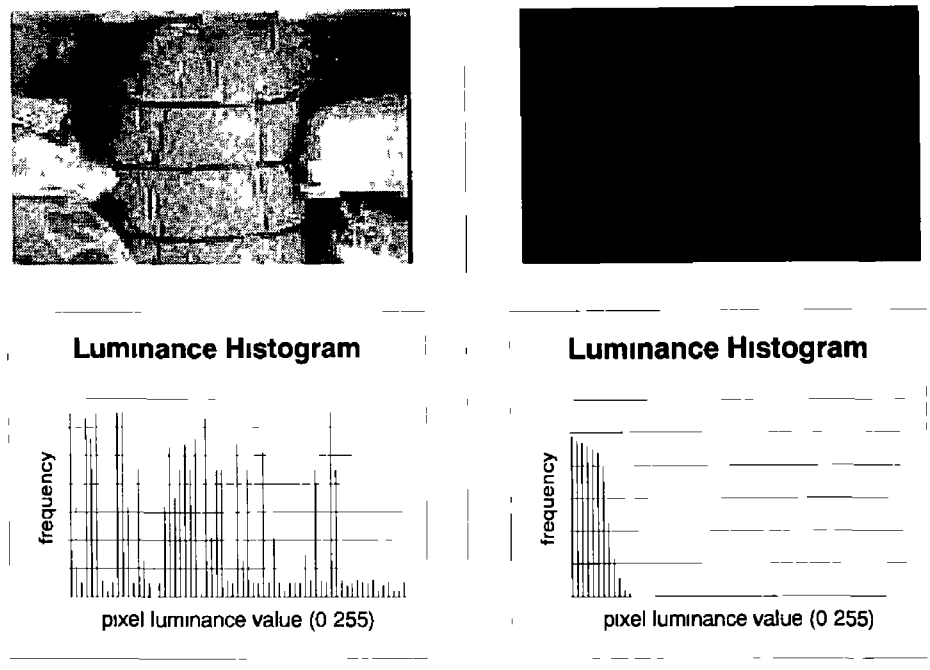


Figure 1-4 Luminance histograms for a generic and a dark video frame

1.3 MPEG-1 BITSTREAM

The criteria described in **Sections 1 2 2 3 & 1 2 2 4** propose a simplistic approach to the task of locating groups of **black** and **silent** video frames within a television programme, which should provide for efficient detection of advertisement breaks However, the method would require direct access to both video pixels and audio samples Therefore it would necessitate a full decode of the captured programme from its compressed format - MPEG-1- which would be highly undesirable from a computational point of view

This thesis proposes that the same assessment and classification of the individual video frames of a captured television programme might be more efficiently, and hence more rapidly, made from the encoded information present in the MPEG-1 bitstream i e it is envisaged that advertisement break detection may be accomplished to the same degree of accuracy as provided by those methods requiring

implementation of a full audio/visual decode, but with the computational burden significantly reduced

2. RELATED WORK

2.1 NOMAD PROJECT (1997)

The *NOMAD* project [2] is a research endeavor administered by Beat Bourquin, Marc Frey and Roger Wetzel, who are members of a Swiss based video processing research group called *Fatal FX*

NOMAD is short for No-More-Advertising The *NOMAD* project is a Silicon Graphics based digital video application that is able to detect if there is an advertisement break on television or not It allows for a VCR to be controlled such that it can record television programmes without advertisements

NOMAD uses the TV station's logo to determine the presence/absence of advertisement breaks The idea is that the TV station logo would disappear during the broadcast of advertisement breaks Most TV station logo's are semi-transparent and therefore don't exhibit constant colours The only information which remains constant are the logo edges, hence the logo detection idea is implemented by an edge detection algorithm

A VCR is connected to a computer and *NOMAD* continously grabs pictures These pictures are then evaluated by *NOMAD* If a logo presence is detected, *NOMAD* will send a start-record command via the serial interface to the VCR (or a stop-record command if no logo is detected)

The *NOMAD* project is based on sound principles which are, according to the authors, fundamental characteristics of broadcast television in continental Europe However, very few of the terrestrial television stations available in Ireland practice the trait of TV channel logo flagging Consequently, this method is challenged when confronted with the characteristics of Irish broadcast television and for this reason, in our case does not justify much further study

[2] Bourquin B , Frey M & Wetzel R., *NOMAD Project* http //www fatalfx com/nomad/

2.2 LIENHART, KUHMUNCH & EFFELSBERG (1997)

Rainer Lienhart, Christoph Kuhmunch and Wolfgang Effelsberg of the University of Mannheim, Germany have also researched into the area of advertisement detection from broadcast television [3] Their study was based on the following principles

- The detection of delimiting black video frames which visually separate advertisement spots from one another, via examination of video pixel data
- Tracking the rate of hard shot cuts throughout the programmes with the expectation of a frequency increase during ad-breaks This was performed by a colour histogram comparison algorithm
- Tracking the action level of individual shots with the expectation of an activity escalation during advertisements This tracking was implemented by edge change ratio and motion vector length indicators

The identification of black video frames was performed by calculating the standard intensity deviation of the pixels in each video frame which was then compared against some threshold The authors report that the detection of the black frame delimiters was a strong and reliable indicator of the presence of advertisement breaks However results showed [3] that the presence of black video frames was not exclusively particular to ad-breaks and that they also infrequently occurred within valuable programme material Hence, this metric could not be solely employed without reinforcement without yielding the detection of numerous false positives which would blemish any accurate ad-break detection success
A high editing frequency is a typical characteristic visually apparent in most ad-breaks The authors quantify the performance of this characteristic in highlighting their presence The method of shot boundary detection via colour histogram comparison was used For each video frame a 64-bin colour histogram, considering only the two most significant bits of each colour band, was computed and then normalised by the number of pixels in the frame

[3] Lienhart, R., Kuhmunch, C. & Effelsberg, W., *On the Detection & Recognition of Television Commercials*, Proc IEEE Conf on Multimedia Computing and Systems pp 509 516 Ottawa Canada 1996

Comparing these on a frame-to-frame basis, the differences were thresholded and if significant, a shot boundary declared The authors report that unaided, this method was successful in locating the presence of ad-breaks but ineffective in locating ad-break boundaries

In attempting to quantify the activity of shots, the authors employed edge change ratio (ECR) and motion vector length (MVL) metrics The advantage of the ECR as a characteristic parameter is that it registers structural changes in a shot such as entering/exiting and moving objects, as well as fast camera operations Experiments showed that the ECR for advertisements was dynamic but was often much more static for the programme content during which they were aired The implementation of both the ECR and MVL metrics provided for excellent ad-break presence/absence detection but with accurate ad-break boundary location proving elusive once again

2.3 THE KARIN PROJECT (2001)

Dan Fry, Eric Hampshire & Thomas Hargrove are three postgraduate students in Santa Clara University, California, who are proficient in the field of digital video analysis Their current venture, which is merely in its design phase, is the development of a windows based digital video application, *KARIN* [4], which would automatically detect the advertisement breaks within pre-recorded cartoon programmes It will be based on the following hypotheses which the authors thoroughly investigate and discuss

- The detection of delimiting black video frames, which visually separate advertisement spots from one another, via examination of video pixel data
- The tracking of the mean audio volume since it is a typical trait for the volume of ad-breaks to be set at a higher volume than that of the episode they are aired during
- Tracking of the TV station logo The idea is that the logo would disappear during the broadcast of advertisement breaks

[4] Fry, D, Hampshire, E, Hargrove, T, *KARIN Automatic Detection of Commercials within Pre-Recorded Cartoon Shows* Preliminary project design report http://www.cse.scu.edu/projects/2000-01/project12/

- Colour separation Since the efforts are to be concentrated on cartoon programmes, the episodes will tend to have fewer colours, spatially further apart in the colour spectrum than those used to synthesise the advertisements
- Frequency of motion Cartoons tend to be static in nature In contrast, advertisements are purposely designed to be snappy and abrupt and typically exhibit a high frequency of motion

Restricting the scope to cartoon based content makes the ad-break detection task much easier In fact it could be assumed that by employing the colour separation metric alone, the objective could be achieved with very high accuracy

The detection of black video frames method is another idea cited As previously discussed, this may realise reliable ad-break detection

It remains to be seen whether the mean audio volume increase during ad-breaks is a generic characteristic of all stations of Irish broadcast television But even if so, due to the occurrence of extended periods of sustained loudness in films and sports programmes, the author believes this metric could be relied upon to 'tweak' estimates but could not be trusted to *solely* generate them

The frequency of motion metric is again particular to the segmentation of cartoon-based content This would be expected to be as accurate an indicator as the colour separation metric but may require more complex computation

Given the standard of previous work in this area, particularly Lienhart *et al.*, it almost seems futile to develop methods for ad-break detection which could only operate on such a limited domain Although they mention the future development of a system that can handle non-cartoon based content involving real people, complex colour schemes and fast movement, the aspirations of this current project suggests that Fry *et al.* are certainly not the leaders in this field

3. MPEG-1 AUDIO/VIDEO STANDARD

3.1 OVERVIEW

The Moving Pictures Experts Group (MPEG), who meet under the International Standards Organisation (ISO), generate international standards for digital video and audio compression MPEG-1 is a standard in five parts

1 ISO/IEC 11172-1 1993

This addresses the problem of combining one or more data streams from the video and audio parts of the MPEG-1 standard with timing information to form a single stream i e multiplexing and synchronisation of audio/video

2 ISO/IEC 11172-2 1993

This specifies a coded representation that can be used for compressing video sequences

3 ISO/IEC 11172-3 1993

This specifies a coded representation that can be used for compressing audio sequences – both mono and stereo

4 ISO/IEC 11172-4 1995

Part 4 specifies how tests can be designed to verify whether bitstreams and decoders meet the requirements as specified in parts 1, 2 and 3

5 ISO/IEC TR 11172-5

Technically not a standard, but a technical report Gives a full software implementation of the first three parts of the MPEG-1 standard

3.2 MPEG-1 VIDEO

3.2.1 Picture Types

3.2.1.1 Intra-Coded Frames (I-frames)

Intra-coded (I) frames are frames that are encoded entirely independently They are not involved in any temporal compression techniques such as motion compensation or motion compensated interframe prediction However, they are encoded in such a way that any spatial redundancy present within the frame is exploited This implies that there is not great efficiency achieved in intra-encoding frames Nevertheless, they are very important since they are used as reference frames for the prediction techniques employed by other frame types Another advantage is that I-frames facilitate random access points

The frequency of occurrence of the I-frames represents a trade-off between compression intensity and error propagation

3.2.1.2 Predicted Frames (P-frames)

As well as exploiting spatial redundancy, Predicted (P) frames are compressed in the temporal domain also, where they rely on data from previous (reference) frames in the sequence This is enhanced by a method called Motion Estimation - a pixel-block matching technique whose objective is to estimate the motion present between reference and current frames and to subsequently 'undo' this estimated motion in order to generate a prediction The predicted frame is then subtracted from the reference frame which yields a prediction residual Given the reference frame, the residual is coupled with the information required to reconstruct the prediction and then the ensemble is encoded accordingly

The combination of spatial and temporal compression achieves a much higher encoding efficiency than that of I-frames It should also be noted that P-frames are allowed to act as reference frames as well as I-frames

3 2 1 3 Bidirectionally Predicted Frames (B-frames)

Bidirectionally predicted (B) frames achieve the highest compression efficiency by employing motion compensation either from a preceding or succeeding frame or both Since they are not used as reference frames, error busy B-frames have no affect any other subsequent frames in the chain
There is no limitation on the number of consecutive B frames that can be used in any group of pictures The encoder makes the decision on how often the different picture types occur and if a good rate of compression is required, then many B-frames will be used

Figure 3-1 illustrates the frame referencing scheme

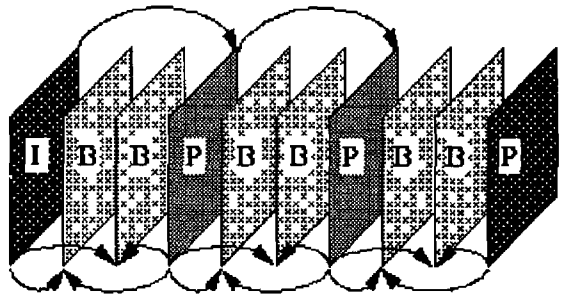


Figure 3-1 Illustration of referencing between frames

3 2 2 MPEG-1 Video Bitstream Hierarchy

The following is a description of the layers of an encoded MPEG-1 bitstream [5] as depicted in **Figure 3-2** [6]

[5] *Video and Audio Compression Website* Dept of Computer Science, Cardiff University Website http://www.cs.cf.ac.uk/Dave/Multimedia/node196.html
[6] Rao K R. and Hwang J J *Techniques & Standards for Image, Video and Audio Coding* Prentice Hall 1996

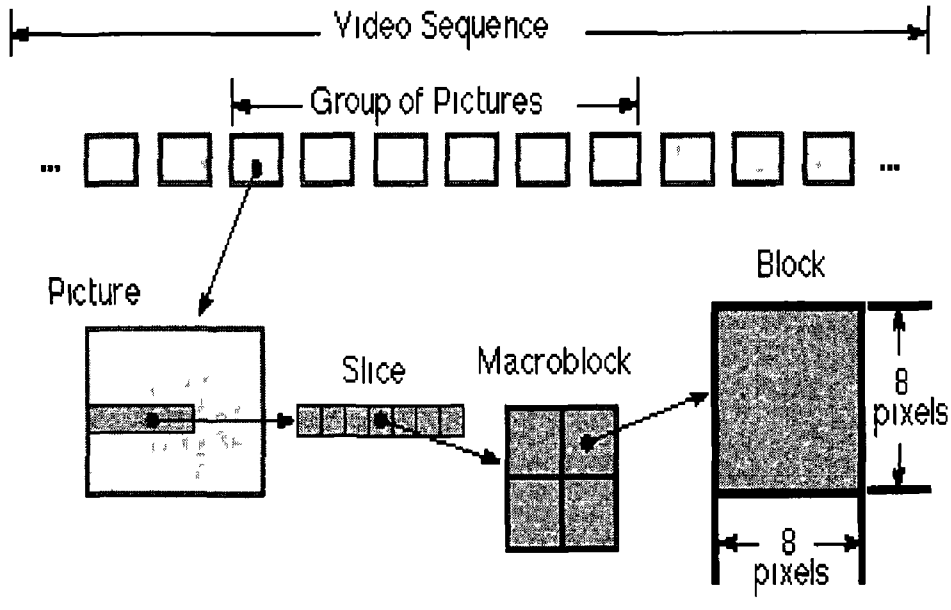


Figure 3-2 The layered MPEG 1 video structure

- The **Video Sequence** provides the video parameters (width, height, pixel aspect ratio and picture rate), bitstream parameters (bit rate, buffer size and constrained parameters flag) and the quantiser default tables

- The **Group of Pictures Layer (GOP)** is a set of pictures that are in continuous display order The overheads provide information on the time code (hrs, mins, seconds, frame) and descriptions of the structure of the GOP A closed GOP represents closed prediction within the group only An open GOP requires decoded pictures from previous GOPs for motion compensation

- At the **Picture Layer** we are told the type of picture - I, P or B (see below) - and its temporal reference We are also given some buffer and encoding information

- Each individual picture is divided into slices The **Slice Layer** provides the information required to partition the data via the vertical position information It also informs us how the quantisation table is scaled in this slice
  The slice layer is important as regards the handling of errors since the decoder can skip any corrupted slice and go to the start of the next slice

- Each slice is subdivided into macroblocks At the **Macroblock Layer**, we are informed of which macroblocks to skip via the address increment information We are also told whether or not the macroblock uses a motion vector (and what type), how the quantisation table is scaled in this macroblock and which blocks are coded ('coded block pattern')

- Each macroblock contains six blocks of (8x8) pixels transformed by a 2D (8x8) Discrete Cosine Transform (DCT) Thus each block consists of a DC-DCT coefficient, which represents its mean luminance intensity, and a number of AC-DCT coefficients, which represent its non-zero frequency content Four of these blocks provide macroblock luminance information, leaving two for chrominance information The **Block Layer** provides information on the weight of the DC-DCT coefficient, the DC difference, the AC-DCT coefficients and an end-of-block (EOB) code The EOB signifies that all DCT coefficients along the zig-zag scan beyond the EOB code are zero See **Figure 3-3**

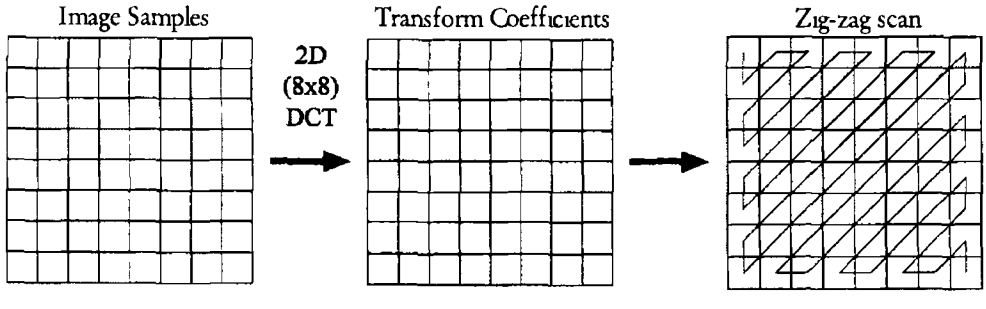


Figure 3 3 Zig zag scanning of 2D (8x8) DCT coefficients

3.3 MPEG-1 LAYER-II AUDIO

Unlike many other audio compression algorithms, which make assumptions about the nature of the audio source, MPEG Audio exploits the perceptual restrictions of the human auditory system, via psychoacoustic weighting of the bit allocation for each frequency subband, to attain its compression

The MPEG Layer-II compression algorithm encodes audio signals as follows the frequency spectrum of the audio signal, bandlimited to 20kHz, is uniformly divided into 32 subbands which approximate the ear's critical bands The subbands are assigned individual bit-allocations according to the audibility of quantisation noise within each subband A pyschoacoustic model of the ear analyses the audio signal and provides this information to the quantiser

Layer-II frames consist of 1152 samples, 3 groups of 12 samples from each of 32 subbands, see **Figure 3-4** A group of 12 samples gets a bit allocation and, if this is non-zero, a scalefactor Scalefactors are weights that scale groups of 12 samples such that they fully use the range of the quantiser (the encoder uses a different scalefactor for each of the three groups of 12 samples within each subband only if necessary)

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 it provides an indication of the maximum power exhibited by any one of the 12 samples within the group

The complete Layer-II data bitstream structure is illustrated in **Figure 3-5**

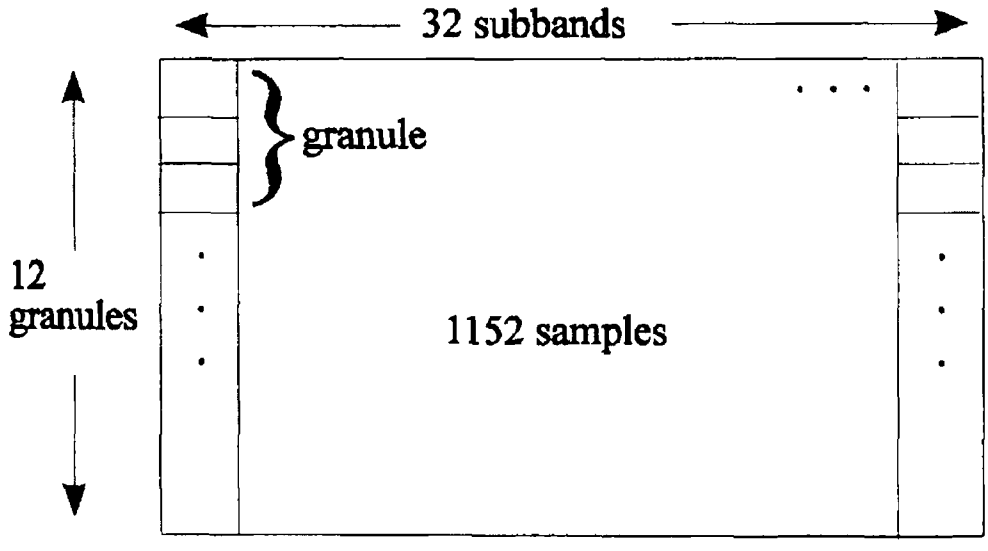


Figure 3-4 Structure of Layer II subband samples

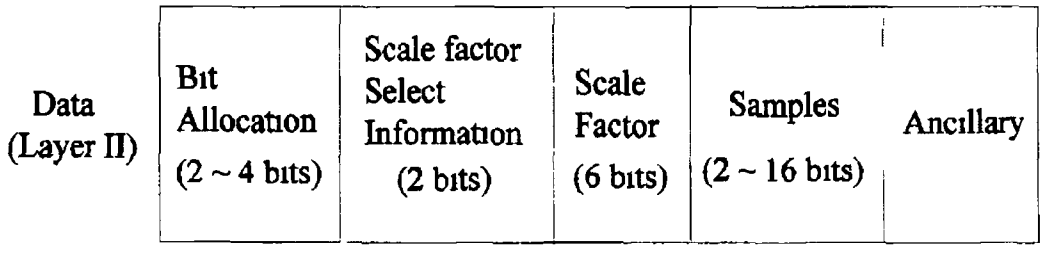


Figure 3 5 The data bitstream structure of Layer-II

4. AD-BREAK DETECTION

4.1 OVERVIEW

As explained in **Chapter-1**, the recognition of **black/silent** video frames, coupled with some pattern recognition algorithm should provide for the efficient detection of advertisement breaks within captured television programmes

The following sections explain how the classification of **black/silent** video frames may be achieved via the examination of data taken directly from the compressed MPEG-1 audio/video bitstream, and also the principles involved in the development of an ad-break pattern recognition scheme The blueprint of the overall ad-break detection scheme is given in **Figure 4-1**

4.2 BLACK VIDEO FRAME DETECTION (DC-DCT)

4.2.1 Frame Luminance Intensity

The four luminance blocks (Y-blocks) present in each macroblock provide the essential information on how dark the macroblock effectively is, hence indicating how dark the overall frame may be

The DC-DCT value of each Y-block represents an average luminance intensity value for that block, to which the visual perception is highly sensitive In contrast, various frequency fluctuations may go unnoticed 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 i e AC variations could be ignored

The proposal was that an average **luminance intensity value** for each video frame could be determined from the DC-DCT coefficients provided within individual Y-blocks which make up the frame This value was expected to be relatively low for dark frames and higher for brighter frames Thus by appropriately thresholding this value, the video frames may be given a **black**/non-**black** catagorisation as desired

In **Graph 4-1**, the frame luminance intensity for all 30000 frames of a random 20 minute television programme, calculated using the above method, is presented as an example

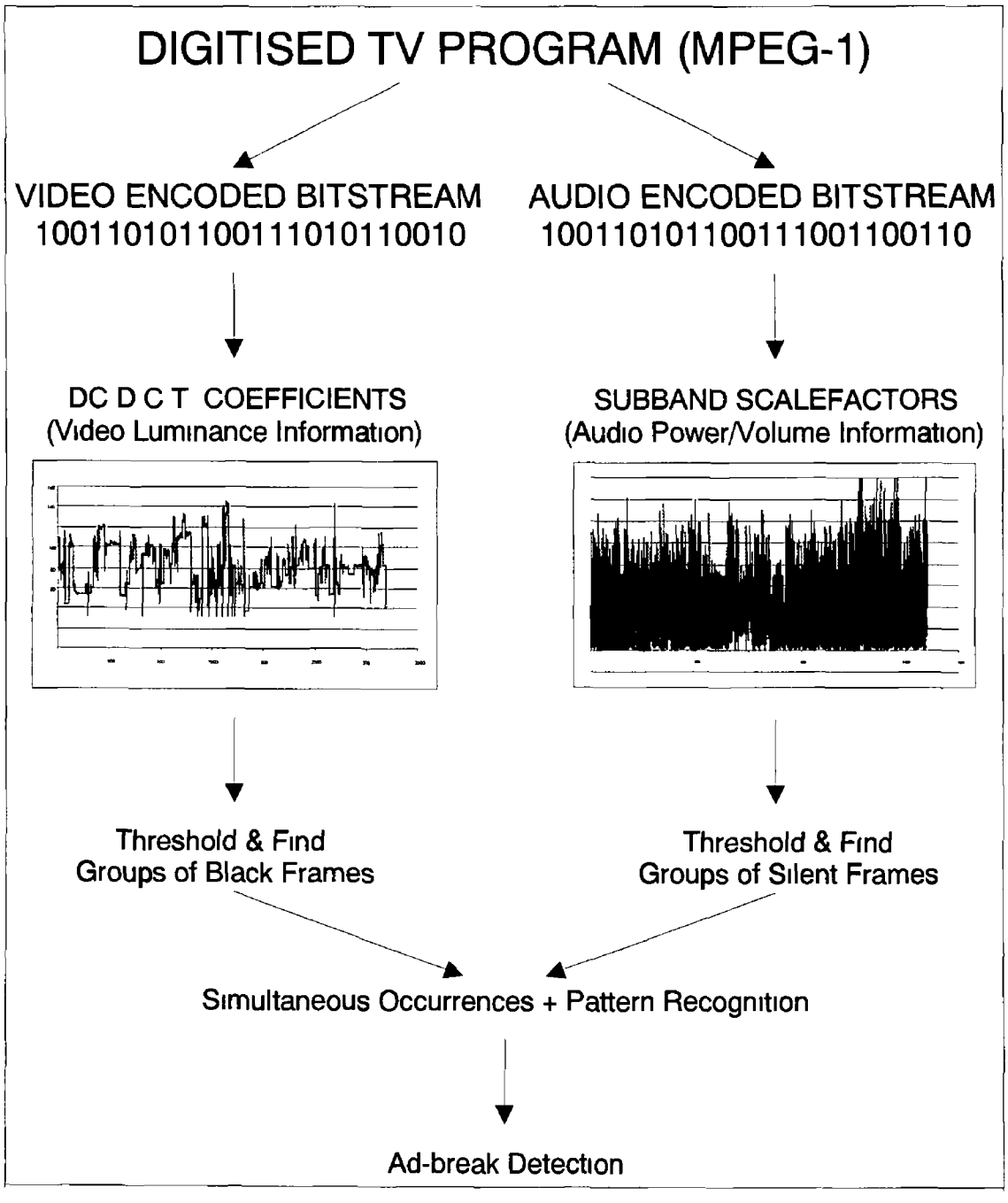
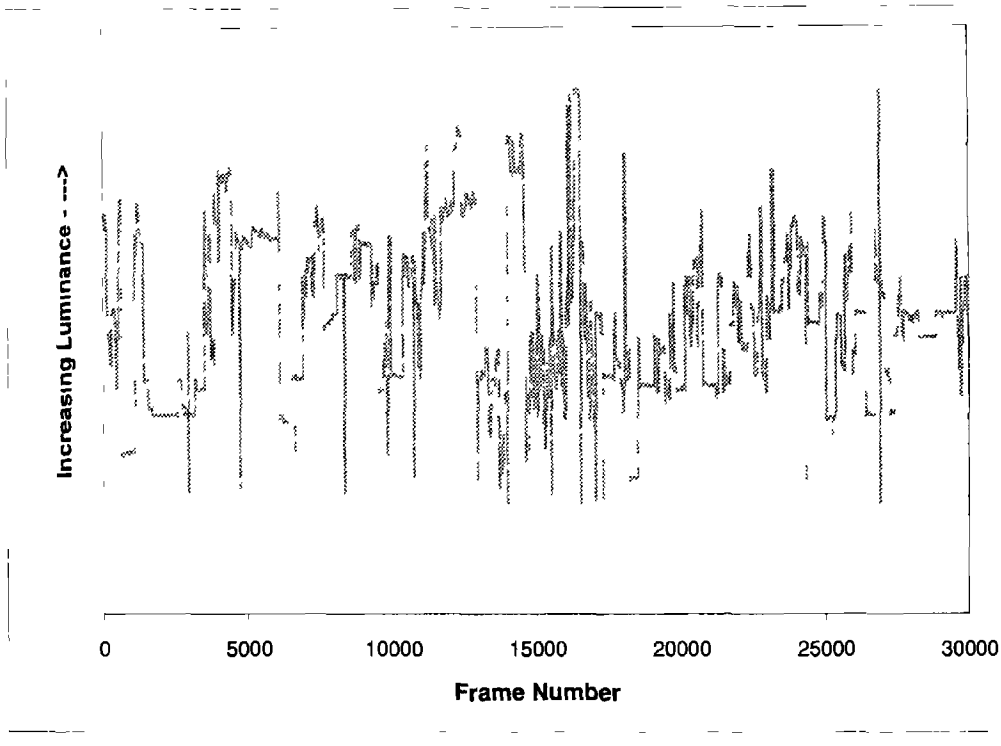


Figure 4-1 Overview of ad-break detection scheme



Graph 4-1 Frame luminance intensity for all frames of a 20 minute programme

4.2.2 Black Frame Threshold

A number of static threshold techniques were investigated However, the following programme-adaptive method provided the most consistency in the results obtained, and was thus chosen as the appropriate scheme

As described in **Section 4 2 1**, each video frame has an associated luminance intensity value, which is derived from the DC-DCT coefficient information A programme-specific mean luminance intensity (**MLI**) value may be obtained by averaging the luminance intensity over all video frames of a programme The **black** frame video threshold (V_{th}) is then defined as some fraction (**X**) of this value

$$V_{th} = X * MLI$$

To determine the optimum value of **X** required for the desired intensity of the thresholding, the following experiments were carried out

Three 20-minute-long television clips were captured [(1), (2) & (3)], and digitally (MPEG-1) encoded by *Físchlár* They each contained a single ad-break and were meticulously chosen such that they exhibited a good mix of both dark and bright video content (e g night and day scenes) The task was to find a global optimum value of **X** which would allow the programme-adaptive threshold technique to distinguish between video frames of a very dark nature and **black** video frames, such as those which delimit advertisement spots, within generic content of varied luminance

Using the DC-DCT coefficient method as described above, a mean luminance intensity value was computed for each of the three clips ($\mathbf{MLI_1}$, $\mathbf{MLI_2}$ & $\mathbf{MLI_3}$) For each clip, numerous values for $\mathbf{V_{th}}$ were calculated by assigning the values [0 025, 0 05, 0 075 0 950, 0 975, 1] to the variable **X**.

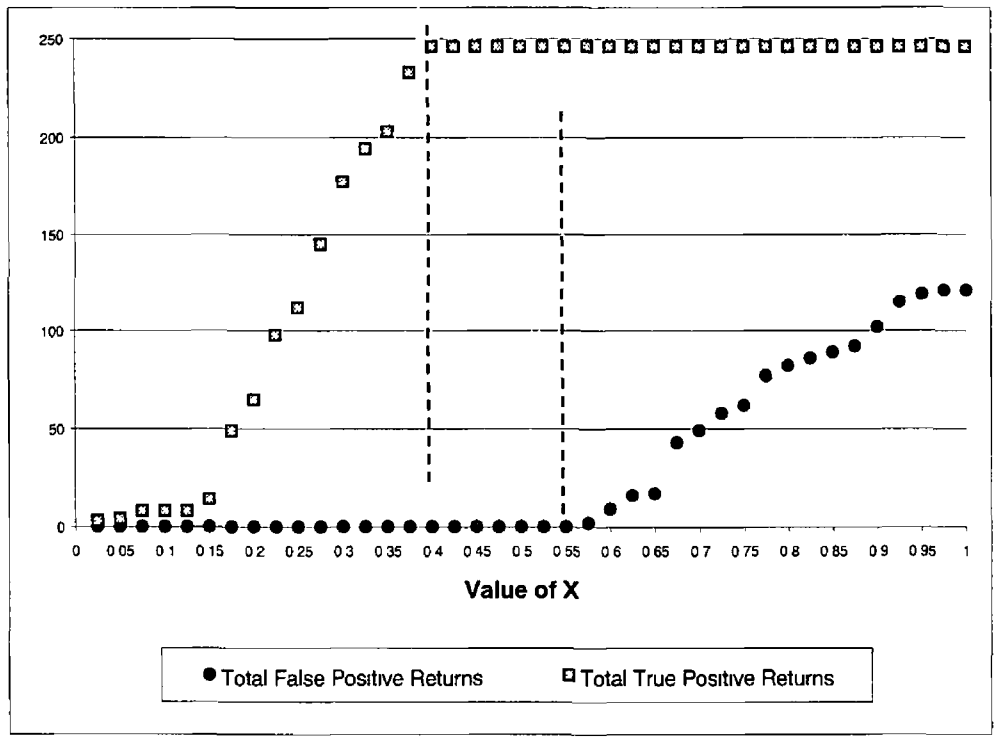
By manual inspection, it was determined that the three ad-breaks consisted of a total of 17 individual advertisement spots delimited by a total of 20 **black/silent** series, amounting to 246 individual **black/silent** frames

Using the same values of **X**, each programme was examined with a different $\mathbf{V_{th}}$ value since they each bore different **MLI** values The detection of **black** video frames for all three programmes was attempted according to the thresholds given by the different **MLI** values and the 40 assumed values of **X**. The results were compared to those of an earlier manual examination

Since the amount of non-ad-break-associated **black** frames present within the programmes had not been quantified, the definition of a *true-positive* result was limited to signify the detection of one of the 246 ad-break-associated **black** video frames and not any other A *false-positive* result represented the mistaken detection of any general non-**black** video frame, irrespective of ad-break (dis-) association The results for all three programmes were accumulated (see **Table 4-1**) and totals illustrated in **Graph 4-2**

X (V_{th})	Clip-1 (MLI_1)		Clip-2 (MLI_2)		Clip-3 (MLI_3)		Total Over All Programmes	
	# True Positives Returned	# False Positives Returned	# True Positives Returned	# False Positives Returned	# True Positives Returned	# False Positives Returned	# True Positives Returned	# False Positives Returned
0 025	0	0	2	0	1	0	3	0
0 05	0	0	2	0	2	0	4	0
0 075	1	0	4	0	3	0	8	0
0 1	1	0	4	0	3	0	8	0
0 125	1	0	4	0	3	0	8	0
0 15	2	0	9	0	3	0	14	0
0 175	15	0	12	0	22	0	49	0
0 2	21	0	19	0	25	0	65	0
0 225	40	0	26	0	32	0	98	0
0 25	48	0	31	0	33	0	112	0
0 275	48	0	46	0	51	0	145	0
0 3	56	0	61	0	60	0	177	0
0 325	69	0	64	0	61	0	194	0
0 35	72	0	64	0	67	0	203	0
0 375	87	0	64	0	82	0	233	0
0 4	87	0	64	0	95	0	246	0
0 425	87	0	64	0	95	0	246	0
0 45	87	0	64	0	95	0	246	0
0 475	87	0	64	0	95	0	246	0
0 5	87	0	64	0	95	0	246	0
0 525	87	0	64	0	95	0	246	0
0 55	87	0	64	0	95	0	246	0
0 575	87	1	64	0	95	1	246	2
0 6	87	1	64	0	95	8	246	9
0 625	87	1	64	0	95	15	246	16
0 65	87	2	64	0	95	15	246	17
0 675	87	2	64	1	95	40	246	43
0 7	87	5	64	4	95	40	246	49
0 725	87	6	64	5	95	47	246	58
0 75	87	8	64	6	95	48	246	62
0 775	87	9	64	10	95	58	246	77
0 8	87	11	64	13	95	58	246	82
0 825	87	12	64	14	95	60	246	86
0 85	87	12	64	17	95	60	246	89
0 875	87	15	64	17	95	60	246	92
0 9	87	22	64	20	95	60	246	102
0 925	87	24	64	31	95	60	246	115
0 95	87	27	64	31	95	61	246	119
0 975	87	28	64	32	95	61	246	121
0 1	87	28	64	32	95	61	246	121

Table 4-1 Results of video threshold experiments



Graph 4 2 Combined results of video threshold experiments

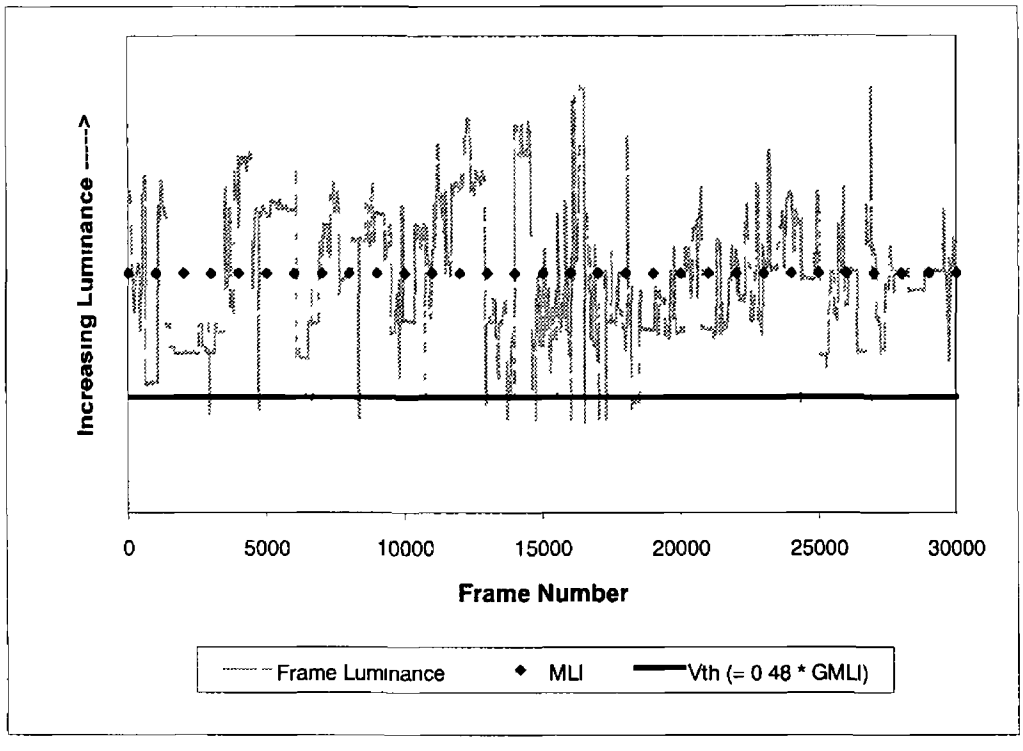
Graph 4-2 indicates that the minimum value for **X** such that all 246 ad-break associated **black** frames were detected is 0 4 However, the maximum value for **X**, such that simply dark frames were not mis-identified as **black** frames is 0 55 These two boundaries of X_{opt} are indicated by the vertical dashed lines in **Graph 4-2** The midpoint of these boundaries is at **X = 0 475** Generalising these results to cover all programmes, it was assumed that the value X_{opt} = 0 48 would suffice to yield **black** frame detection to the same degree of accuracy for any further generic video content examples

$$X_{opt} = 0\ 48$$

Thus the final definition of V_{th} for black frame detection for a general TV programme with mean luminance intensity **MLI** is

$$V_{th} = 0\ 48 * MLI \qquad \textbf{Equation 4-1}$$

Graph 4-3 is the same example of frame luminance given in Section 4 2 1 (**Graph 4-1**), however, the programme mean luminance intensity (**MLI**) and corresponding **black** frame threshold values (V_{th}) are now also illustrated



Graph 4 3 Frame luminance intensity for all frames of a 20 minute programme Also shown is the mean luminance intensity and its corresponding **black** frame threshold

4.3 SILENT VIDEO FRAME DETECTION (SCALEFACTORS)

4 3.1 Audio Volume Level

4 3 1 1 Scalefactor Superposition

The subband scalefactor of a group of 12 samples effectively indicates the maximum power exhibited by any one sample within such a group Thus it provides a means by which a variable audio power level may be tracked on a per-12-sample basis without necessitating a decode from the compressed bitstream

The proposal was that an **audio volume level** for each video frame could be determined by a superposition of the scalefactors corresponding to the groups of audio samples to which the video frames are temporally associated - see **Figure 4-2** This volume level was expected to remain significantly low for **silent** video frames and to be high for video frames associated with more substantial audio content Thus by thresholding this value, the video frames may be assigned a **silent**/ non-**silent** catagorisation as desired

For demonstration purposes, in **Graph 4-4**, the associated audio volume level for the first 300 video frames of the 20 minute example clip from **Section 2 4 1** is plotted

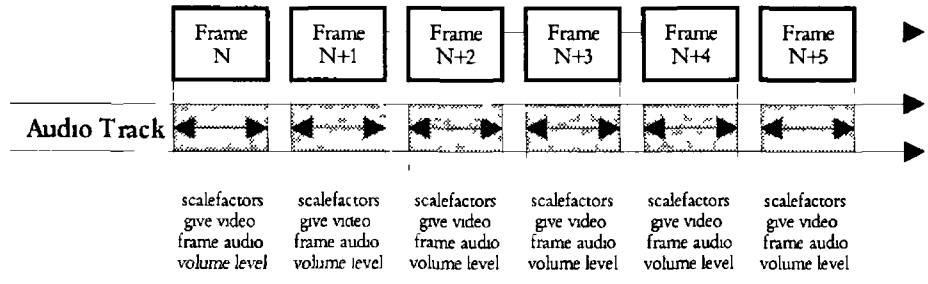
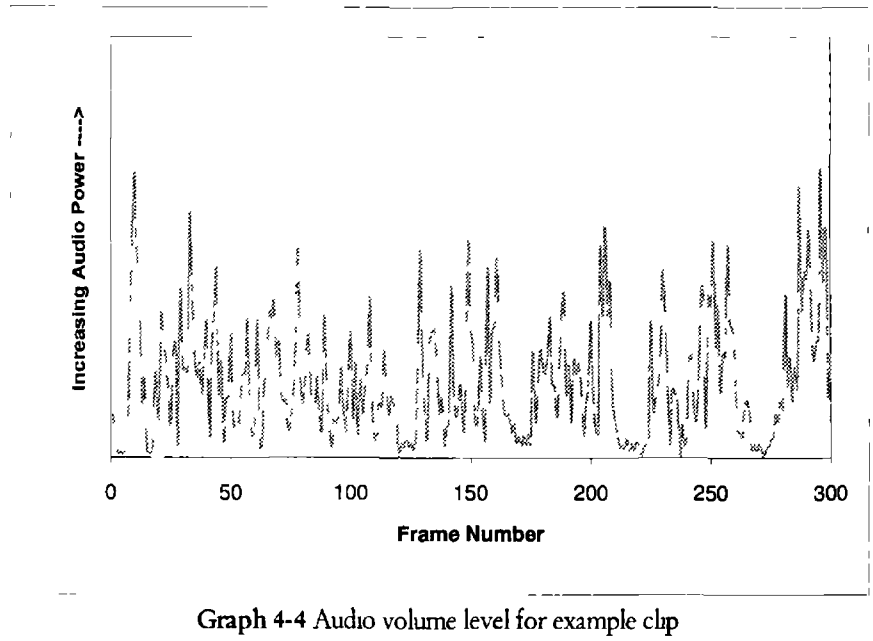


Figure 4 2 Video frame audio volume levels generated from scalefactors corresponding to temporally associated audio

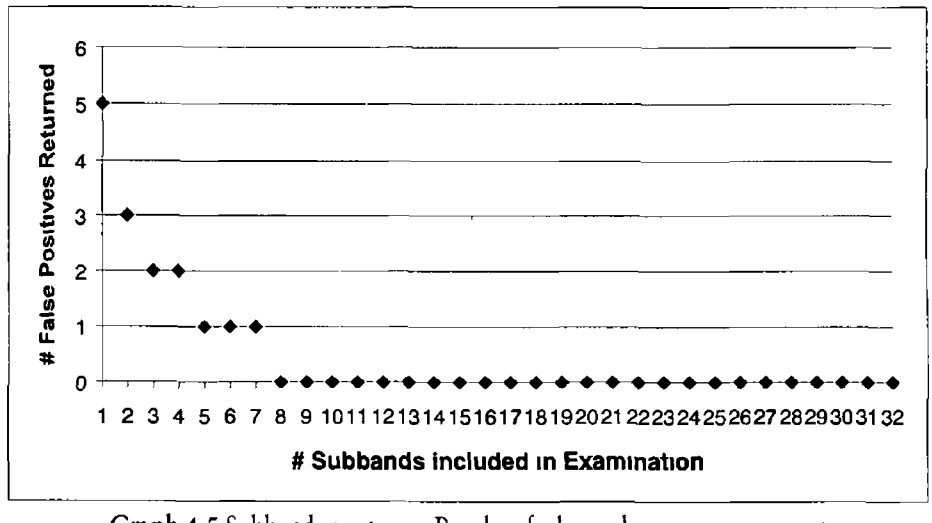


Graph 4-4 Audio volume level for example clip

4 3 1 2 Reduction of Cut-off Frequency

Since it is typical of an audio signal to have most of its energy corresponding to relatively low frequencies, it was expected that examination of just the bottom end of the frequency spectrum would provide sufficient information for which the **silent** video frame catagorisation could still be accurately made To investigate this proposal, an MPEG-1 Layer-II volume varying audio track was examined which contained 28 **silences** - aurally identified with minimum duration of 1 second Initially the scalefactors from subbands 1-32 of the bitstream were inspected, and detected **silences** noted The number of subbands providing scalefactor information was then successively reduced by one (i e scalefactors from subbands 1-31 were examined, then subbands 1-30 etc) until just the first subband's scalefactors were inspected exclusively For all cases the scalefactor information was used to detect the number of silences (greater than 1 second in duration) contained within the audio track.

The results showed that for every single case as described above, all 28 true silences were detected as expected However **Graph 4-5** describes how the rate of *false-positive* detection increased as the subband cutoff was reduced



Graph 4-5 Subband rejection – Results of silence detection experiment

Graph 4-5 shows that the number of false positive returns remained at zero until merely subbands 1-7 were examined - one false positive was then returned i e 29 silences detected As the subband cutoff was further reduced, the number of false positive returns increased steadily It was shown that when just the first subband was examined, the scalefactor information yielded the detection of 33 silences in the audio signal, corresponding to 5 false positive returns

Thus, for this particular audio example, it was shown that the minimum requirement for accurate results in a **silent** frame detection task was to examine the scalefactors from the first 8 subbands inclusively It was expected that this could be extended to more generic audio signals without inaccuaracies However, to allow for a margin of possible error, the subband cut-off was slightly rounded up to include subbands 9 & 10

Thus, the **silent** frame location section of the ad-break detection algorithm was performed by examination of scalefactors from subbands 1-10 inclusively The maximum frequency encoded by the MPEG-1 Layer-II standard is 20kHz, thus the first 10 subbands corresponds to an examination of the energy present from 0-6kHz

4.3.2 Silent Frame Threshold

A programme-adaptive thresholding technique, similar to that used for the **black** frame detection method described in **Section 4 2 2**, was employed for the audio classification

As mentioned in **Section 4 3 1**, each video frame has an associated audio volume level, which is derived from the scalefactors from subbands 1-10 of the encoded bitstream

A programme-specific mean audio volume (**MAV**) value may be obtained by averaging the audio volume levels over all video frames of a programme The **silent** frame audio threshold (A_{th}) is then defined as some fraction (**Y**) of this value

$$A_{th} = Y \cdot MAV$$

To determine the optimum value of **Y** required for the desired intensity of the thresholding, the following experiments were carried out

In addition to the three clips used in Section 4 2 2 [(1), (2) & (3)], three further clips were captured and digitally (MPEG-1) encoded by *Físchlár* These clips [(4), (5) & (6)] were meticulously chosen such that they exhibited a good mix of loud and quiet moments while noticeably never going completely **silent**. i e there was a continuous audio signal presence for the entire duration of the clips The task was to find an optimum (global) value of **Y** which would, allow the programme-adaptive threshold technique to distinguish between video frames associated with non-**silent** but low-volume audio, particularly present in clips (4), (5) & (6), and **silent** video frames present in clips (1), (2) & (3)

Using the subband scalefactor method as described above, a mean audio volume value was computed for all six captured clips (MAV_1, MAV_2, MAV_3, MAV_4, MAV_5 & MAV_6)

In each case, numerous values for A_{th} were calculated by assigning the values [0 0605, 0 0610, 0 0615, 0 0620 0 1005, 0 1010, 0 1015] to the variable **Y**

It was known from the manual inspection performed in **Section 4 2 2**, that the combined number of delimiting ad-break-associated **silent** frames within clips (1), (2) & (3) amounted to 246

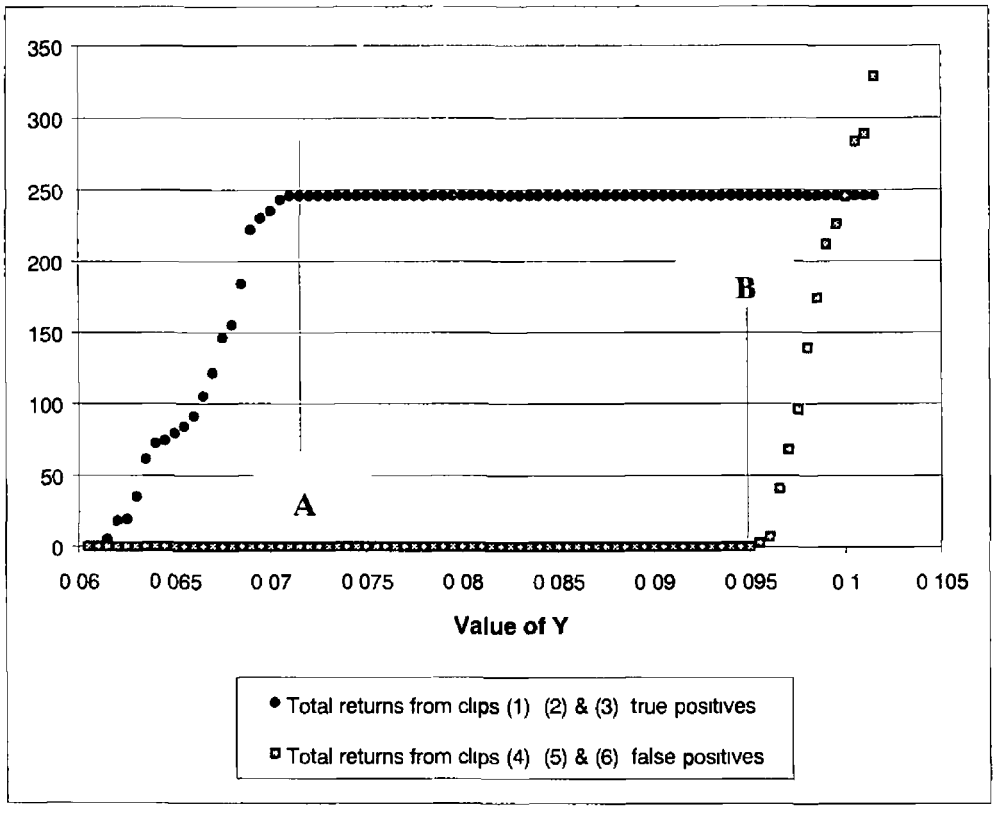
Using the same values of **Y**, each programme was examined with a different A_{th} value since they each bore different **MAV** values The detection of **silent** video frames for all six programmes was attempted according to the thresholds given by the different **MAV** values and the assumed values of **Y**

Since the amount of non-ad-break-associated **silent** frames present within clips (1), (2) & (3) had not been quantified, the definition of a *true-positive* result was limited to signify the detection of one of the 246 ad-break-associated **silent** video frames and not any other A *false-positive* result represented the mistaken detection of any general non-**silent** video frame, irrespective of ad-break (dis-) association, within clips (4), (5) & (6) The results for all six programmes were accumulated (see **Table 4-2**) and totals illustrated in **Graph 4-6**

Y (A_{th})	# True Positives Returned from Clip-1 (MAV_1)	# True Positives Returned from Clip-2 (MAV_2)	# True Positives Returned from Clip-3 (MAV_3)	# False Positives Returned from Clip-4 (MAV_4)	# False Positives Returned from Clip-5 (MAV_5)	# False Positives Returned from Clip-6 (MAV_6)	**Totals** # True Positives Returned from Clips (1), (2) & (3)	**Totals** # False Positives Returned from Clips (4), (5) & (6)
0 0605	0	0	0	0	0	0	0	0
0 061	0	0	0	0	0	0	0	0
0 0615	0	1	4	0	0	0	5	0
0 062	0	8	10	0	0	0	18	0
0 0625	0	9	10	0	0	0	19	0
0 063	0	19	16	0	0	0	35	0
0 0635	16	20	25	0	0	0	61	0
0 064	19	21	32	0	0	0	72	0
0 0645	19	21	34	0	0	0	74	0
0 065	20	22	37	0	0	0	79	0
0 0655	20	26	38	0	0	0	84	0
0 066	23	30	38	0	0	0	91	0
0 0665	26	34	45	0	0	0	105	0
0 067	32	44	45	0	0	0	121	0
0 0675	33	51	62	0	0	0	146	0
0 068	33	53	69	0	0	0	155	0
0 0685	49	61	74	0	0	0	184	0
0 069	75	64	83	0	0	0	222	0
0 0695	80	64	91	0	0	0	230	0
0 07	86	64	95	0	0	0	235	0
0 0705	87	64	95	0	0	0	243	0
0 071	87	64	95	0	0	0	246	0
0 0715	87	64	95	0	0	0	246	0
0 072	87	64	95	0	0	0	246	0
0 0725	87	64	95	0	0	0	246	0
0 073	87	64	95	0	0	0	246	0
0 0735	87	64	95	0	0	0	246	0
0 074	87	64	95	0	0	0	246	0
0 0745	87	64	95	0	0	0	246	0
0 075	87	64	95	0	0	0	246	0
0 0755	87	64	95	0	0	0	246	0
0 076	87	64	95	0	0	0	246	0
0 0765	87	64	95	0	0	0	246	0
0 077	87	64	95	0	0	0	246	0
0 0775	87	64	95	0	0	0	246	0
0 078	87	64	95	0	0	0	246	0
0 0785	87	64	95	0	0	0	246	0
0 079	87	64	95	0	0	0	246	0
0 0795	87	64	95	0	0	0	246	0
0 08	87	64	95	0	0	0	246	0
0 0805	87	64	95	0	0	0	246	0
0 081	87	64	95	0	0	0	246	0

0 0815	87	64	95	0	0	0	246	0
0 082	87	64	95	0	0	0	246	0
0 0825	87	64	95	0	0	0	246	0
0 083	87	64	95	0	0	0	246	0
0 0835	87	64	95	0	0	0	246	0
0 084	87	64	95	0	0	0	246	0
0 0845	87	64	95	0	0	0	246	0
0 085	87	64	95	0	0	0	246	0
0 0855	87	64	95	0	0	0	246	0
0 086	87	64	95	0	0	0	246	0
0 0865	87	64	95	0	0	0	246	0
0 087	87	64	95	0	0	0	246	0
0 0875	87	64	95	0	0	0	246	0
0 088	87	64	95	0	0	0	246	0
0 0885	87	64	95	0	0	0	246	0
0 089	87	64	95	0	0	0	246	0
0 0895	87	64	95	0	0	0	246	0
0 09	87	64	95	0	0	0	246	0
0 0905	87	64	95	0	0	0	246	0
0 091	87	64	95	0	0	0	246	0
0 0915	87	64	95	0	0	0	246	0
0 092	87	64	95	0	0	0	246	0
0 0925	87	64	95	0	0	0	246	0
0 093	87	64	95	0	0	0	246	0
0 0935	87	64	95	0	0	0	246	0
0 094	87	64	95	0	0	0	246	0
0 0945	87	64	95	0	0	0	246	0
0 095	87	64	95	0	0	0	246	0
0 0955	87	64	95	3	0	0	246	3
0 096	87	64	95	5	1	1	246	7
0 0965	87	64	95	24	6	11	246	41
0 097	87	64	95	31	15	22	246	68
0 0975	87	64	95	41	18	37	246	96
0 098	87	64	95	54	35	50	246	139
0 0985	87	64	95	63	37	74	246	174
0 099	87	64	95	77	49	86	246	212
0 0995	87	64	95	77	62	87	246	226
0 1	87	64	95	87	69	90	246	246
0 1005	87	64	95	96	81	107	246	284
0 101	87	64	95	97	85	107	246	289
0 1015	87	64	95	112	101	116	246	329

Table 4 2 Results of audio threshold experiments



Graph 4 6 Combined results of audio threshold experiments

Graph 4-6 shows that the minimum value for **Y** such that all 246 ad-break-associated **silent** frames are detected from clips (1), (2) and (3) is 0 071 However, the maximum value for **Y**, such that simply quiet frames are not mis-identified as **silent** frames in clips (4), (5) & (6) is 0 095 These two boundaries of Y_{opt} are indicated by the vertical dashed lines (**A** & **B**) in **Graph 4-6**

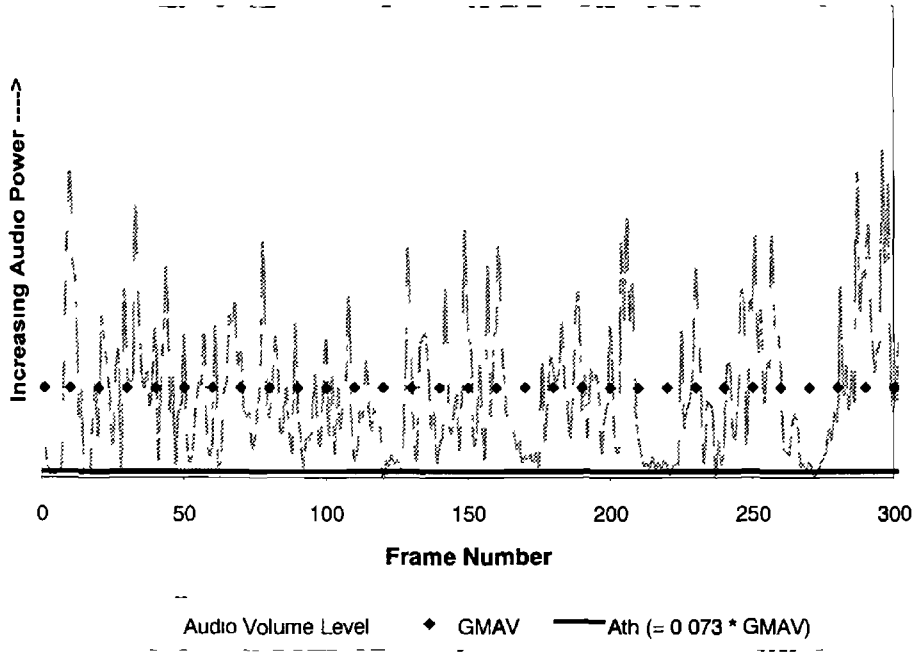
In generalizing these results, it was expected that the lower boundary (**A**) would remain reasonably constant for all ad-breaks from all television broadcasts, while the existing position of the upper boundary (**B**) could not be expected to generalise well as it would tend to vary much more on a programme to programme basis For these reasons, Y_{opt} was biased such that it resided close to the lower boundary (**A**) at 0 073

$$Y_{opt} = 0\ 073$$

Thus the final definition of A_{th} for **silent** frame detection for a programme with mean audio volume **MAV** is

$$A_{th} = 0.073 * MAV \qquad \text{Equation 4-2}$$

Graph 4-7 is the same example of audio volume level given in **Section 4.3.1** (**Graph 4-4**), however, the programme mean audio volume (**MAV**) and **silent** video frame threshold values (A_{th}) are now also shown



Graph 4-7 Audio volume level for example clip. Also shown is the mean audio volume and its corresponding **silent** frame threshold

4.4 Pattern Recognition For Ad-Breaks

As explained, the occurrence of **black/silent** video frame series may indicate the existence of an ad-break. However, it is possible, and maybe quite probable, that these indicators also occur during the valuable material of 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 possible consequence of mis-identification of valuable programme content, some strict conditions have to be enforced

4.4.1 Ad-break Pattern Characteristics

Over 20 different ad-breaks were examined from the broadcasts of several television stations It was observed that all ad-breaks consistently exhibited the following characteristics - see **Appendix-A**

- The delimiting **black** and **silent** frame series between individual advertisement spots had an average duration of 10 frames The shortest duration lasted just 8 frames
- The average length of all advertisement spots was 25 seconds The longest duration of any single spot was 76 seconds
- The average number of spots per ad-break was 7 The least amount of spots in one single ad-break was found to be 4

Although relating to just the 20 ad-breaks examined, these characteristics were assumed to be generically consistent for all ad-breaks broadcast by the television networks in Ireland

4.4.2 Ad-break Pattern Assumptions

The three ad-break characteristics provided a basis on which the following assumptions about any typical ad-break structure could be made See **Figure 4-3**

- Any delimiting **black/silent** frame series was assumed to have duration of at least 6 video frames
- Any individual advertisement spot was assumed to be always less than 90 seconds in duration

- Any particular ad-break was assumed to contain at least 3 individual spots

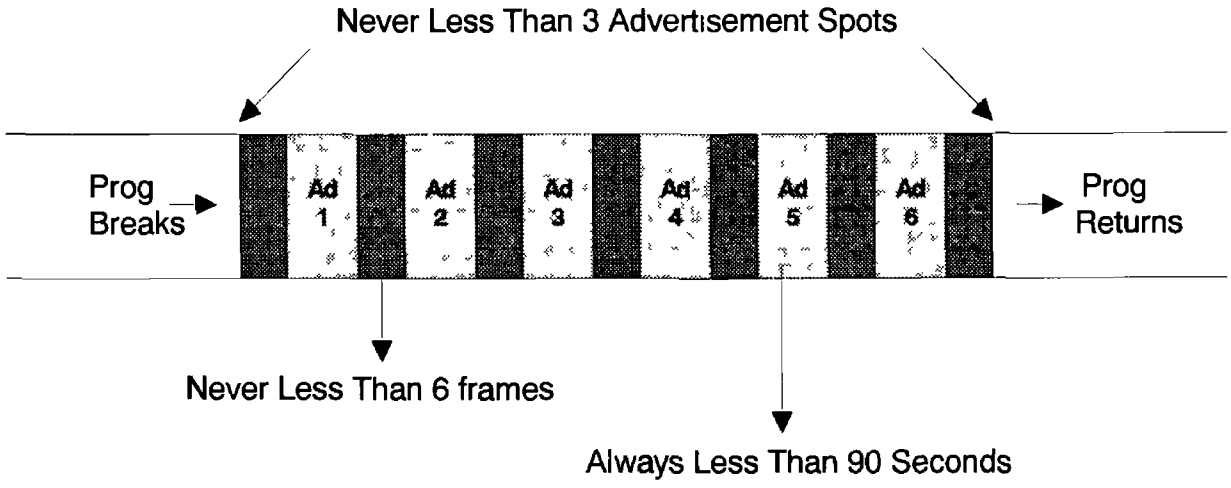


Figure 4-3 Ad-break pattern assumptions

4.4.3 Ad-break Pattern Criteria

Once the task of locating all the **black/silent** video frame series within the content of a captured television programme has been completed, the duration, frequency and relative locations of the series are examined and compared against the following ad-break recognition criteria, which were derived from the ad-break structure assumptions

1 Any detected **black/silent** frame series consisting of at least 6 consecutive frames would continue to be recognised while the others would be subsequently disregarded (This aims to initially reject any diminutive rogue **black/silent** frame series which may innocuously occur within the valuable programme content)

2 It was assumed that the maximum spot length is 90 seconds If the minimum number of spots comprising any ad-break is 3, this implies the presence of at least 4 delimiting **black/silent** frame series Consequently, only series who

possess at least three neighbouring series, within respective 90sec windows of each other continue to be recognised Others would be disregarded

It is the combined effect of the above clauses which is expected to provide the success in accurately preventing mis-recognition of programme content for advertisement material **Figure 4-4** explains how detected **black/silent** video frame series within a television programme are interpreted and subsequently how ad-breaks are recognised

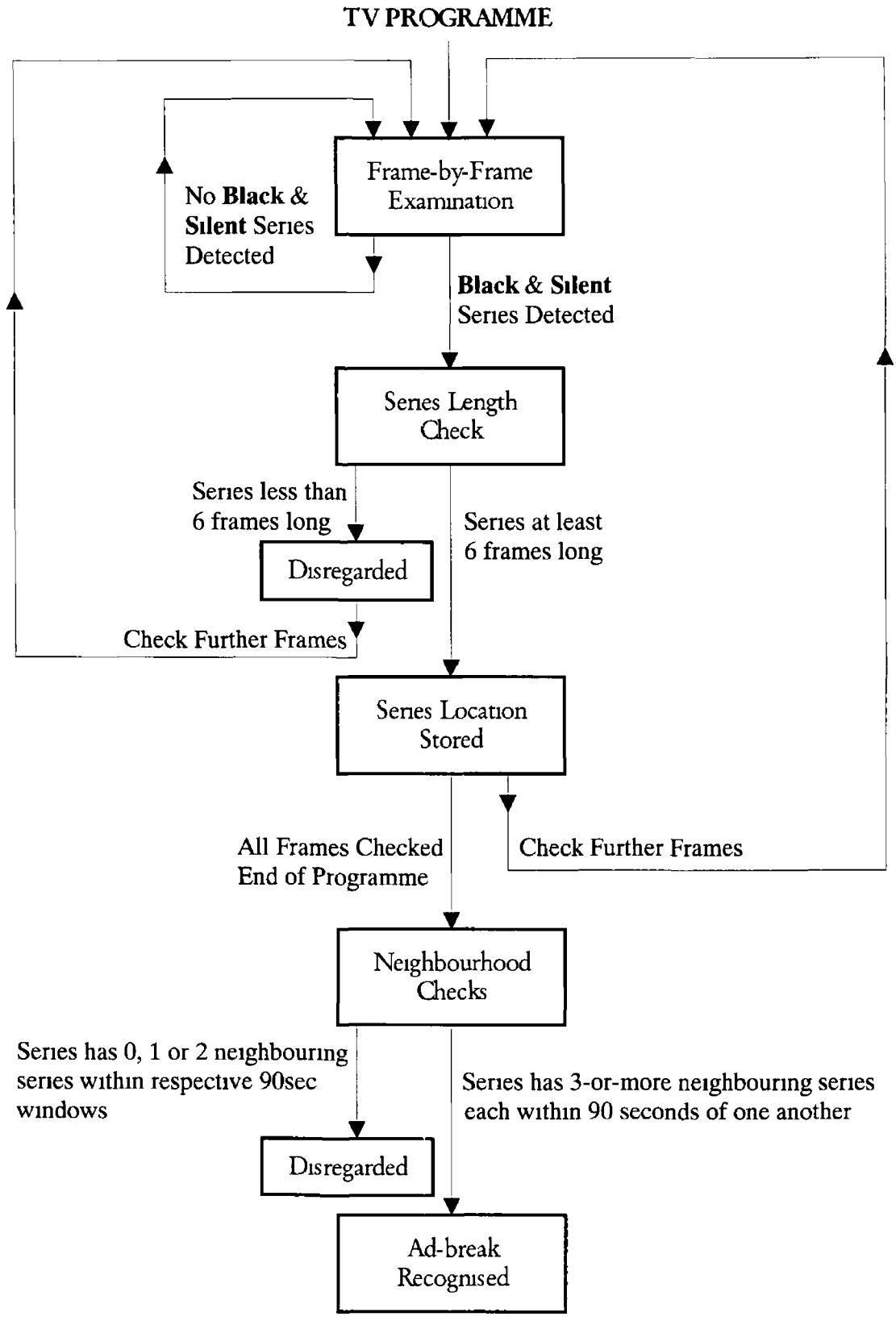


Figure 4-4 Implementation of criteria for the pattern recognition of ad breaks

5. ILLUSTRATION, ANALYSIS & RESULTS

5.1 ILLUSTRATION

The following illustration describes the complete process of ad-break detection within a 20-minute (30000-frame) television programme example It begins with the detection of **black/silent** video frames from the compressed MPEG-1 bitstream and then explains how the ad-break pattern criteria are invoked and how they contribute to pinpointing the precise location of the ad-breaks within a broadcast television programme

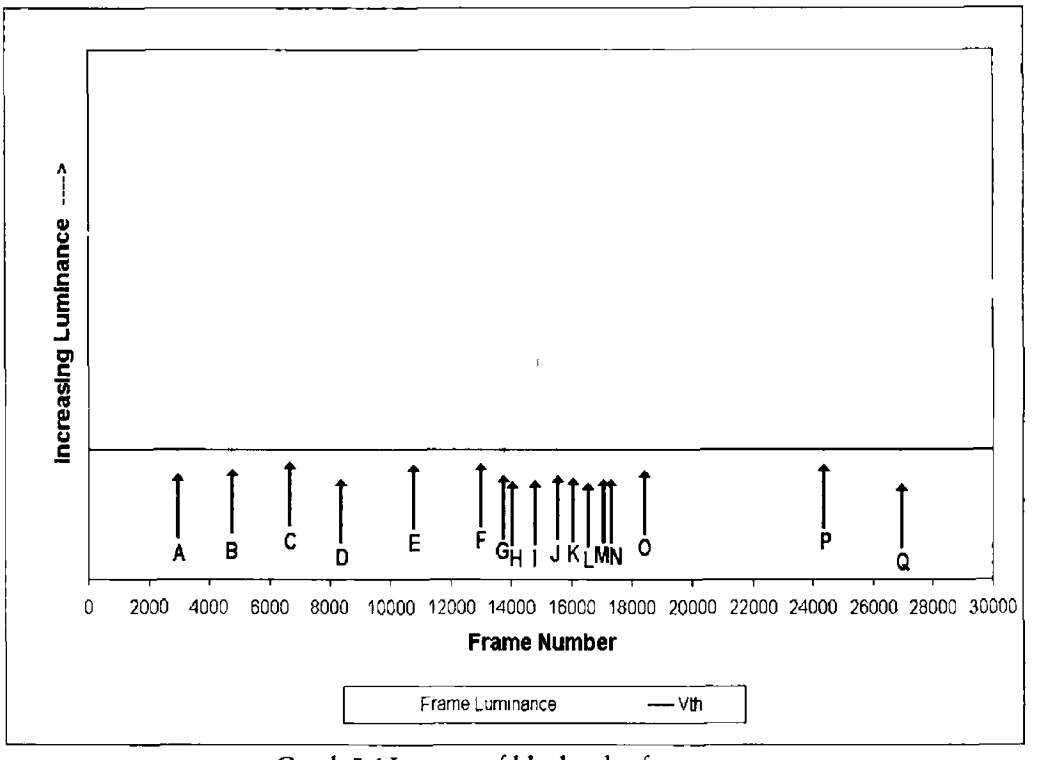
5.1.1 Video Examination

- The DC-DCT coefficients of each Y-block of each video frame were stripped from the MPEG-1 video bitstream of the captured programme and used to obtain frame luminance intensity values
- A mean luminance intensity (**MLI**) value was calculated by averaging all frame luminance intensity values over the entire clip
- The **black**-frame threshold (V_{th}) was expressed as the value corresponding to 48% of the **MLI** value [see **Section 4 2 2** especially **Equation 4-1**]

Graph 5-1 describes how the frame luminance intensity varied over the course of the programme V_{th} crossings correspond to **black**-frame-series and are indicated by the arrows labeled A-Q

5.1.2 Audio Examination

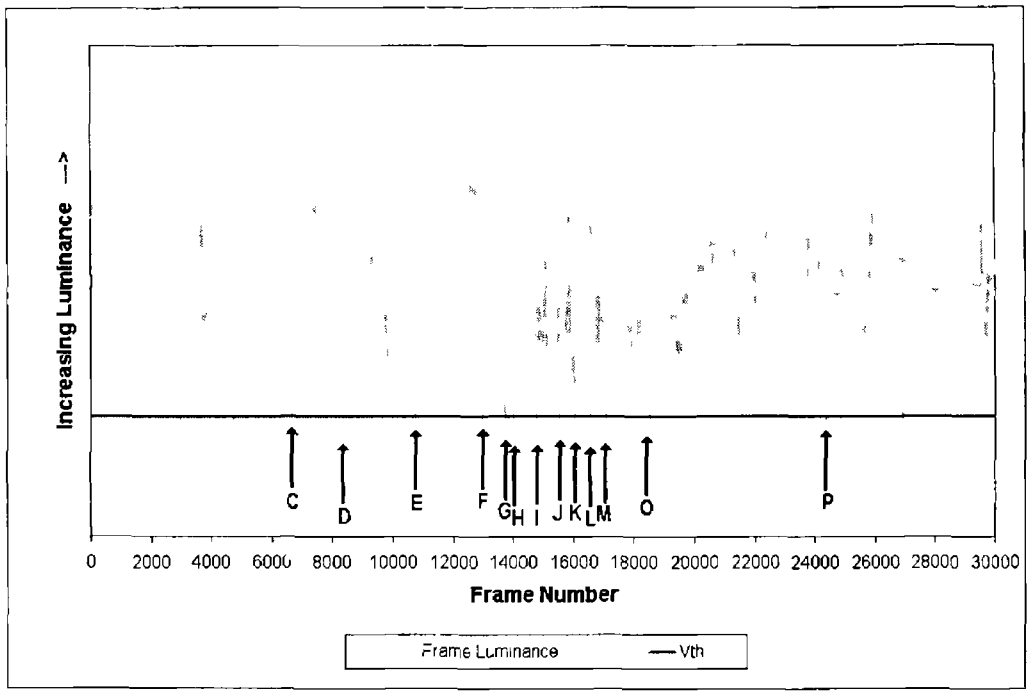
- The scalefactors corresponding to subbands 1-10 of the Layer-II encoded audio signal were stripped from the bitstream and manipulated to obtain video-frame audio volume levels



Graph 5 1 Location of **black** video frame series

- A mean audio volume (MAV) value was calculated by averaging all audio volume levels over the entire clip
- The **silent**-frame threshold (A_{th}) was expressed as the value corresponding to 7 3% of the MAV value [see **Section 4 3 2** especially **Equation 4-2**]

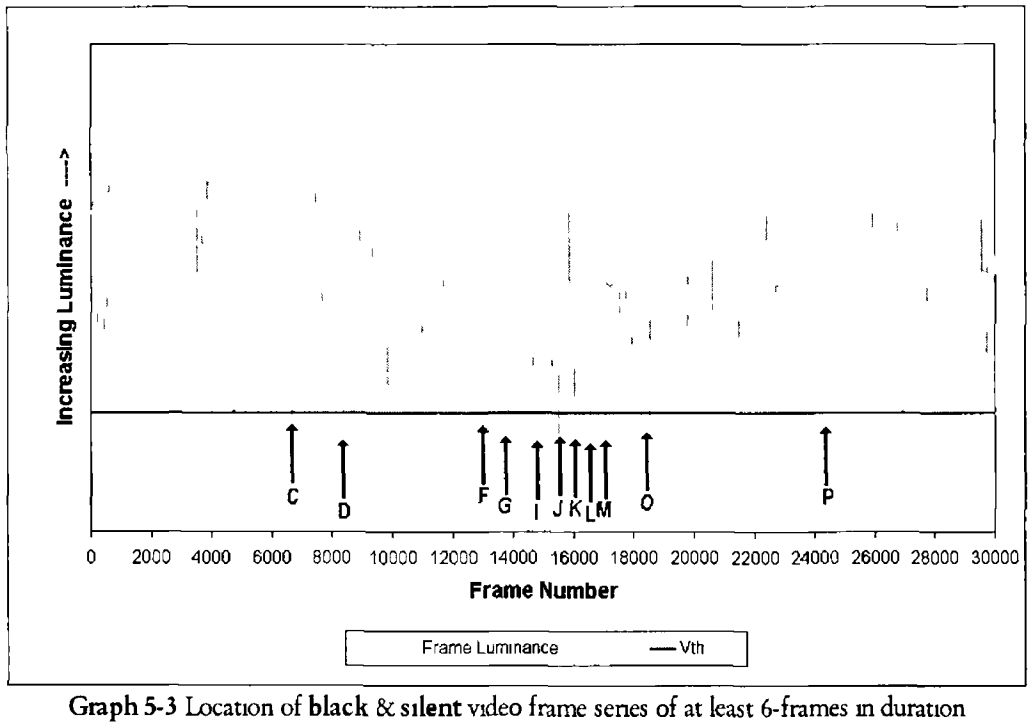
The results of the **silent**-frame detection process suggested that the **black** frames in series A, B, N & Q in **Graph 5-1** were <u>not</u> also **silent** and as a result were ignored **Graph 5-2** is similar to **Graph 5-1** but with the arrows removed from the abovementioned series



Graph 5-2 Location of **black** & **silent** video frame series

5.1.3 Ad-break Pattern Recognition

The first of the pattern criteria stated that the minimum permissible length of an ad-break-associated **black/silent** video frame series was 6-frames Investigation showed that series E & H did not meet this criteria and as a result were ignored **Graph 5-3** is similar to **Graph 5-2** but with the arrows removed from the abovementioned series The remaining criteria stated that only series who possess at least three neighbouring series, within respective 90-second (2250-frame) windows of each other continue to be recognised

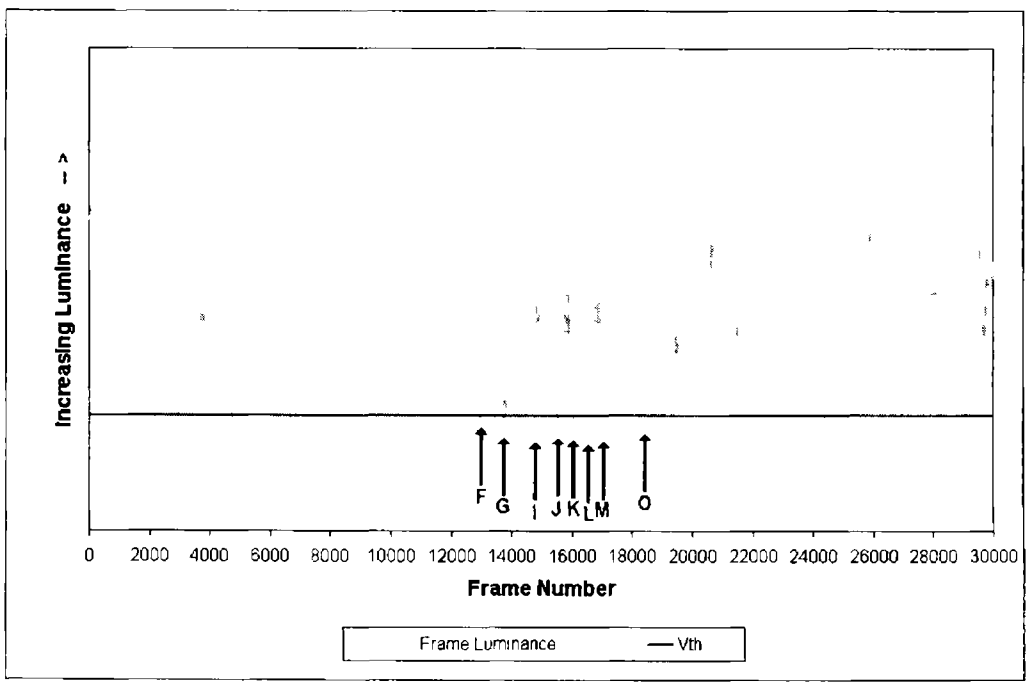


Graph 5-3 Location of **black** & **silent** video frame series of at least 6-frames in duration

- Series-C was ignored since it only possessed one neighbour within 2250 frames [Series-D]
- Series-D was ignored since it only possessed one neighbour within 2250 frames [Series-C]
- Series-F was retained since it possessed at least three neighbours within 2250-frame windows of each other [Series-G, -I & -J]
- Series-G was retained since it possessed at least three neighbours within 2250-frame windows of each other [Series-F, -I & -J (or others)]
- Series-I was retained since it possessed at least three neighbours within 2250-frame windows of each other [Series-G, -F & -J (or others)]
- Series-J was retained since it possessed at least three neighbours within 2250-frame windows of each other [Series-I, -K & -L (or others)]

- Series-K was retained since it possessed at least three neighbours within 2250-frame windows of each other [Series-L, -M & -O (or others)]
- Series-L was retained since it possessed at least three neighbours within 2250-frame windows of each other [Series-M, -O & -K (or others)]
- Series-M was retained since it possessed at least three neighbours within 2250-frame windows of each other [Series-L, -K & -O (or others)]
- Series-O was retained since it possessed at least three neighbours within 2250-frame windows of each other [Series-M, -L & -K]
- Series-P was ignored since it possessed no neighbours within 2250 frames

Graph 5-4 is similar to **Graph 5-3** but with the arrows removed from the series which did not meet the abovementioned criteria



Graph 5-4 Location of **black** & **silent** video frame series which meet all the ad-break pattern criteria

5.1 4 Final Decision: Recognised Series

The definition of a **recognised black/silent** video frame series is one which meets all requirements of the ad-break pattern criteria **Recognised** series flag the presence of ad-breaks within television programmes In this case, series F, G, I, J, K, L, M & O in **Graph 5-4** are such series They suggested the location of an ad-break beginning with series-F and finishing after series-O, within the given content
Comparison of this result with that of a manual investigation showed that the system was 100% accurate in detecting the location of the entire ad-break from within the programme

5.2 ANALYSIS & RESULTS

5.2.1 Test Corpus

Físchlár provided for the digital capture of 17 television programmes from 4 different channels [labeled (A), (B), (C) & (D)] and encoded them according to the MPEG-1 format The recordings were meticulously chosen such that all exhibited significant content diversity and all but two contained one complete ad-break somewhere in the middle
The ad-break detection scheme procedures, as illustrated in **Section 5 1,** were executed on all 17 clips Evaluation of the system was performed by comparison of results achieved against a manual record of the true location of the ad-breaks within each clip, to the nearest second Results are given in **Table 5-1**

5 2.2 Results

The absence of falsely identified content is evident from the results in **Table 5-1** In clip **6 (B),** the detected end of the ad-break is at 770 seconds, but the true end is later, at 790 seconds This is recorded as 20 missed seconds of ad-break material
The total number of seconds in the true ad-break (= 790 – 603) = 187 seconds

The total number of seconds detected as ad-break (= 770 – 603) = 167 seconds

A superposition of these quantities over all 17 clips was performed to give four overall values, which are illustrated in **Figure 5-1**

Clip No (channel)	Programme content description	Length of clip (sec)	# Detected black/silent video frame series	# Recognised black/silent video frame series	Suggested ad-break location in clip (second-second)	True ad-break location in clip (second second)
1 (A)	Chat Show	600	5	5	193s – 281s	193s – 281s
2 (A)	News Broadcast	1200	4	4	468s – 516s	468s – 538s
3 (A)	Music Show	600	5	4	258s – 347s	257s – 347s
4 (B)	News Broadcast	1200	11	11	113s – 351s	113s – 351s
5 (B)	Soap Opera	1200	7	7	779s – 935s	779s – 935s
6 (B)	Sports Show	1200	11	8	603s – 770s	603s – 790s
7 (B)	Cookery Show	2400	13	10	1102s – 1335s	1102s – 1335s
8 (C)	Sports Show	1800	10	10	773s – 919s	773s – 919s
9 (C)	Youth Magazine	1800	14	7	185s – 272s	185s – 272s
10 (C)	Game Show	1500	7	5	659s – 772s	659s – 772s
11 (C)	Comedy Quiz	1800	6	6	774s – 913s	774s – 934s
12 (D)	News Broadcast	1800	9	9	1415s – 1660s	1415s – 1660s
13 (D)	Cartoon	1800	12	11	798s – 975s	798s – 975s
14 (D)	Music Show	1800	10	8	688s – 853s	688s – 853s
15 (D)	Nature Show	2400	13	9	1357s – 1558s	1357s – 1558s
16 (A)	News Broadcast	1200	1	0	None Detected	None
17 (B)	Sci-fi Show	2700	4	0	None Detected	None

Table 5 1 Results of ad-break detection experiments

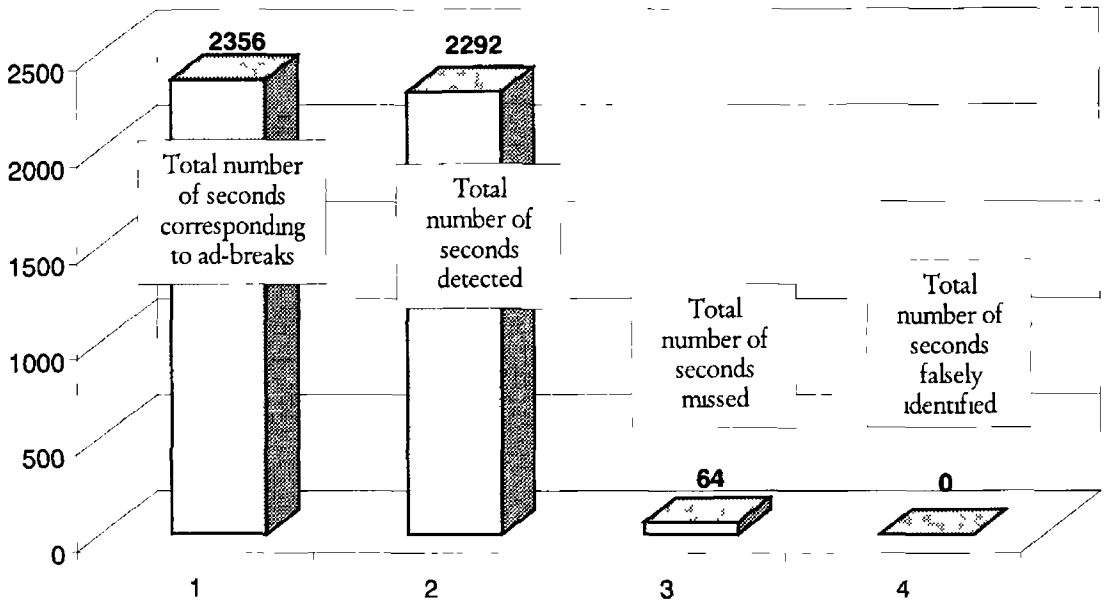


Figure 5 1 Total number of seconds corresponding to ad-breaks, detected ad-breaks, missed ad breaks and falsely identified ad breaks

5 2 2 1 Precision & Recall

For a better insight into the individual accuracy the result for each clip, two important figures of merit for the ad-break detection system were calculated

- The **Recall** measure is a value representing the percentage of all detected material corresponding to true ad-breaks

$$\text{Recall} = \frac{100 * [\text{Length of ad-break (seconds)} - \text{No seconds missed}]}{\text{Length of ad-break (seconds)}}$$

- The **Precision** measure is a percentage showing how accurate the system is at exclusively detecting ad-break material

$$\text{Precision} = \frac{100 * [\text{Length of ad-break (seconds)} - \text{No seconds missed}]}{\text{Length of ad-break (seconds)} - \text{No seconds missed} + \text{No seconds falsely identified}}$$

Example Precision & Recall for Clip 6 (B)

For clip **6 (B)** the Recall and Precision figures were calculated as follows (from the information in **Table 5-1**)

Length of ad-break (= 790 – 603) = 187 seconds

No seconds falsely identified = 0 seconds

No seconds missed = 20 seconds (ad-break-end detected 20 seconds early)

Precision = 100 [(187 – 20) / (187 – 20 + 0)] = **100**

Recall = 100 [(187 – 20) / 187] = **89 3**

Results following similar calculations performed on all clips are presented in **Table 5-2**

Clip	Precision	Recall
1 (A) - Chat Show	100	100
2 (A) - News Broadcast	100	68 6
3 (A) - Music Show	100	98 8
4 (B) - News Broadcast	100	100
5 (B) - Soap Opera	100	100
6 (B) - Sports Show	100	89 3
7 (B) – Cookery Show	100	100
8 (C) - Sports Show	100	100
9 (C) - Youth Magazine	100	100
10 (C) - Game Show	100	100
11 (C) – Comedy Quiz	100	86 9
12 (D) – News Broadcast	100	100
13 (D) - Cartoon	100	100
14 (D) – Music Show	100	100
15 (D) – Nature Show	100	100
16 (A) – News Broadcast	N/A	N/A
17 (B) – Sci-fi Show	N/A	N/A

Table 5-2 Precision & Recall values for all clips

6. CONCLUSIONS & FURTHER WORK

6.1 SYSTEM EVALUATION

In all, 17 clips comprising 450 minutes of digital video were analysed 15 of the 17 clips contained a single ad-break. 2 clips contained no ad-breaks such that system resilience against mis-identification of content may be further tested The following are the main points to be noted

- The system detected the occurrence of all 15 ad-breaks
- Not once did the system falsely detect an ad-break.
- All ad-break detections attained a Precision percentage of 100%
- 11-out-of-15 of the ad-break detections attained a Recall percentage greater then 98%

Clips **2** (**A**), **6** (**B**) & **11** (**C**) performed relatively poorly compared to all others It was noted that their common downfall was the ad-break-end-boundary being detected prematurely Manual investigation revealed that the reason for this was that for the ad-breaks in all three clips, the final delimiter (occurring between final advertisement spot and return of programme) did not meet the threshold length of **6** frames, thus violating a section of the ad-break pattern criteria imposed in the detection process Consequently, the system did not detect the occurrence of the final spot within the ad-break, but identified the ad-break-end-boundary with the end of the last **recognised** delimiter (which actually corresponded to the end of the penultimate spot)

Numbers of individual advertisement spots comprising overall ad-breaks were counted for the three clips It was then concluded that the ad-break detection experiments for clips **6** (**B**) and **11** (**C**) resulted in 1-out-of-7 and 1-out-of-6 spots missed respectively, thus giving Recall percentages of **89 3** and **86 9** However, clip **2** (**A**) News Broadcast' resulted in 1-out-of-4 spots missed within its ad-break, which represented non-detection of a much larger proportion of the overall ad-break. Hence, this clip only attained a meagre Recall percentage of **68 6**

For this application, high Precision at the expense of Recall represents a favourable bias The consistently high Precision figures indicated that the ad-break pattern criteria achieved high success in the prevention of mis-recognition of valuable programme content This is desirable since from a *Físchlár* point of view, a user may reasonably endure the system failing to detect the odd advertisement spot now and again without much fuss, but would not tolerate the skipping of relevant content

6.2 FURTHER WORK

6 2.1 Ad-break Pattern Criteria Optimisation

The system succeeded very well except on 3-out-of-15 occasions [Clips 2 (A), 6 (B) & 11 (C)] when the conditions incorporated to combat false identification of valuable programme content actually prevented the detection of the final advertisement spot within their respective ad-breaks Clearly the current ad-break pattern criteria have biased the system slightly towards the precision end of the spectrum For this reason, any future work could possibly involve further optimisation of the trade-off between consistently precise ad-break detection and the prevention of false identification It remains to be seen if by altering the existing thresholds of the current pattern conditions the existing few rogue recall values may be improved, without a general compromise in precision, which would not be tolerated Alternatively, entirely different conditions may be derived to replace any which are shown to be flawed

6.2.2 System Speed

All the abovementioned experiments were performed on a P3 700MHz machine running the Linux Red Hat 7 2 operating system The times taken to perform the ad-break detection procedures for all of the clips are given in **Appendix-B** It is clear from these results that, on average, the system performs the ad-break detection task in a time corresponding to approximately 6% of the programme length Clearly, the process is not very time consuming and computationally moderate This can be

attributed to the method of bitstream processing i e instead of performing a complete MPEG-1 decode and subsequently working with large quantities of raw audio/visual data, the system exploits features of the encoded bitstream, which are akin to average values of the raw data These values convey the same required information as the raw data but with much smaller bulk.

However, it was envisaged that the processing times could still be improved by spatially sub-sampling in the video domain and temporal sub-sampling in the audio domain

In **Section 4 2 1** it was described how frame luminance intensity values were generated by manipulating the DC-DCT values of all Y-blocks of the video frames of a programme It would be interesting to see if equally accurate **black** frame detection could be achieved if the luminance intensity values were generated from the DC-DCT coefficients of just a limited number of Y-blocks instead See **Figure 6-1** By doing this, the computation overheads would be reduced in intensity, hence providing for a faster processing time For example by just selecting every other Y-block for contribution, the computation intensity for the generation overall frame luminance values would be halved It is assumed that this shortcut would not bear significant negative consequence for the accuracy of **black** frame detection It remains to be shown how selective the spatial sampling may be before significant difficulties in discriminating **black** frames becomes apparent

In **Section 4 3 1** it was described how video-frame audio-volume-values were generated by manipulating the first 10 subband's scalefactor values corresponding to the audio signal associated to the entire temporal duration of a video frame (see **Figure 4-2**) Assuming a frame rate of 25fps, the temporal duration of a video frame corresponds to 0 04 seconds So in generating frame audio volume values, 0 04 seconds of audio data has to be processed per video frame It would be interesting to see if equally accurate **silent** video frame detection could be achieved if the frame audio volume values were generated from shorter temporal windows instead of that corresponding to entire frame duration By doing this, the computation overheads would be reduced, hence providing for a faster processing time **Figure 6-2** and **Figure 6-3** suggest some ways in which this may be accomplished

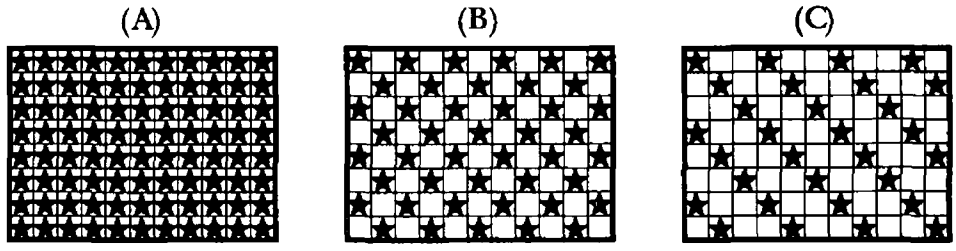


Figure 6-1 (A) All Y-blocks contribute to frame luminance value (B) Every other Y-block contributes to frame luminance value (C) Every third Y-block contributes to frame luminance value

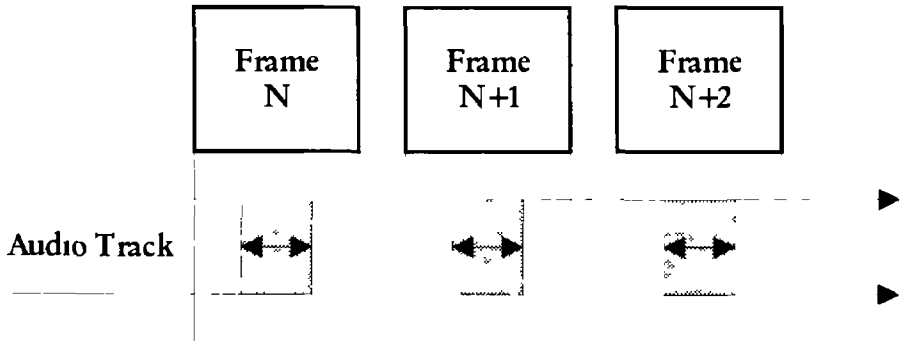


Figure 6 2 Alternate method for generation of video-frame audio-volume-level values

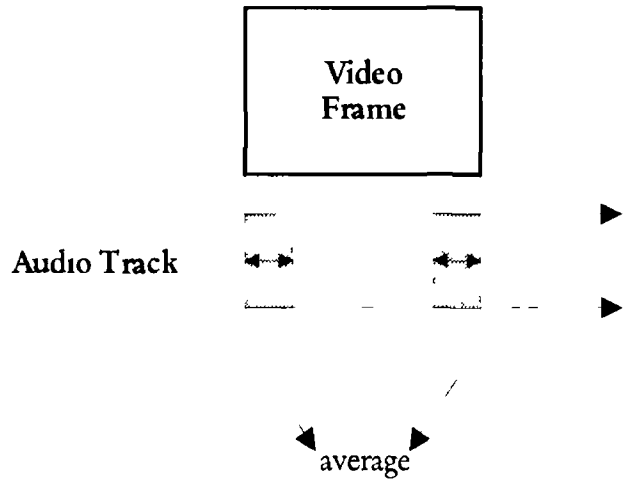


Figure 6 3 Alternate method for generation of video-frame audio-volume-level values

Figure 6-2 shows how the frame audio volume levels may be generated by examining the audio signal once per video frame in relatively small windows temporally centered half-way through the frame duration In **Figure 6-3** the audio signal is examined twice per video frame, in smaller windows at the beginning and end of the frame duration The frame audio levels are then derived by averaging these two sections By employing either of these alternate methods, the computation intensity for the generation of overall frame audio volumes would be reduced Again, it is assumed that either of these shortcuts would not significantly alter the accuracy of **silent** frame detection It remains to be shown how much more selective the temporal sub-sampling may become before significant difficulties in discriminating **silence** becomes apparent

6.2.3 Alternate Techniques

The approach taken in this work was to sift through the MPEG-1 representation of the audio/visual content of an encoded television programme in pursuit of **silent** & **black** video frame series which, pending a criteria probe, may highlight the location of an advertisement break.

Other methods such as those described in **Chapter-2** may be employed to produce detection results of a similar accuracy, however most seem to require largely complex computation and thus would require much longer processing times

In introducing this subject, it was mentioned that most advertising television stations feature the characteristic of **black** and **silent** frame delimiters separating individual advertisement spots Clearly, for any television stations whose ad-breaks do not exhibit this trait, alternate methods are required For any such cases, it is proposed to deploy the shot-cut rate tracking technique as described in **Section 2 2** It is envisaged that this, coupled with some timing aspects (e g broadcast time is usually sold in discrete units), could provide sufficient information enabling an alternative method for automatic advertisement/programme differentiation with results that may rival that of the discussed method

6.2.4 Foreign Broadcasting Formats

The system has been designed to suit the ad-break format characteristics of Irish broadcast television as described in **Section 4 4 1** The formats are common to all the available commercial television stations broadcast in Ireland

To configure the system to operate on foreign broadcasts with different advertising formats, one must re-develop the procedures mentioned in **Sections 4 4 1, 4 4 2 & 4 4 3** accordingly, where pattern characteristics are recorded, pattern assumptions made and then criteria derived

BIBLIOGRAPHY

1 Lee H., Smeaton A., O'Toole C., Murphy N., Marlow S & O'Connor N., *The Físchlár Digital Video Recording, Analysis, and Browsing System*, Proc Content based Multimedia Information Access (RIAO 2000), Vol 2, pp 1390-1399, Paris, France, 12-14 April 2000

2 Carroll, D., *Advertisement Detection*, Internal technical report, Dublin City University 2000

3 *The Offical MPEG Committee Website* http://www.cselt.it/mpeg/

4 Marshall D., *Video and Audio Compression Website* http://www.cs.cf.ac.uk/Dave/Multimedia/node196.html

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

6 Bourquin B., Frey M. & Wetzel R., *NOMAD Project* http://www.fatalfx.com/nomad/

7 Lienhart, R., Kuhmunch, C. & Effelsberg, W., *On the Detection & Recognition of Television Commercials*, Proc IEEE Conf on Multimedia Computing and Systems, pp 509-516, Ottawa, Canada, 1996

8 Fry, D., Hampshire, E., Hargrove, T., *Automatic Detection of Commercials within Pre-Recorded Cartoon Shows* Preliminary project design report http://www.cse.scu.edu/projects/2000_01/project12/

9 Li, Y & Jay Kuo, C.-C., *Detecting the Commercial Breaks in Real TV Programs Based on Audiovisual Information*, Proc SPIE Vol 4210, Internet Multimedia Management Systems 2000, pp 225-236

APPENDICES

APPENDIX-A

The characteristics of over 20 ad-breaks were examined from the broadcasts of several different television channels The findings are tabulated below

Ad-break (Channel)	# Spots in Ad-break	Longest Spot Length	Average Spot Length	Shortest Black/ Silent Frame Series Duration	Average Black/ Silent Frame Series Duration
1 (A)	5	48 seconds	24 seconds	11 frames	14 frames
2 (A)	6	65 seconds	32 seconds	8 frames	9 frames
3 (A)	5	39 seconds	35 seconds	10 frames	11 frames
4 (A)	11	76 seconds	26 seconds	9 frames	10 frames
5 (A)	9	54 seconds	19 seconds	10 frames	12 frames
6 (B)	5	65 seconds	31 seconds	8 frames	9 frames
7 (B)	7	55 seconds	21 seconds	9 frames	12 frames
8 (B)	4	64 seconds	20 seconds	8 frames	8 frames
9 (B)	5	32 seconds	23 seconds	12 frames	13 frames
10 (B)	9	51 seconds	28 seconds	8 frames	10 frames
11 (C)	5	43 seconds	25 seconds	9 frames	9 frames
12 (C)	9	26 seconds	15 seconds	9 frames	11 frames
13 (C)	5	64 seconds	30 seconds	10 frames	11 frames
14 (C)	10	68 seconds	22 seconds	8 frames	10 frames
15 (C)	5	38 seconds	25 seconds	9 frames	9 frames
16 (D)	8	51 seconds	27 seconds	9 frames	10 frames
17 (D)	5	61 seconds	16 seconds	10 frames	11 frames
18 (D)	8	37 seconds	24 seconds	9 frames	10 frames
19 (D)	9	39 seconds	29 seconds	8 frames	9 frames
20 (D)	8	49 seconds	20 seconds	10 frames	12 frames
21 (D)	11	61 seconds	27 seconds	8 frames	9 frames
TOTAL	149	-------	-------	-------	-------
AVERAGE	7 09	-------	**24 7 seconds**	-------	**10 4 frames**

APPENDIX-B

All ad-break experiments were carried out on a P3 700MHz machine running the Linux Red Hat 7 2 operating system The times taken to perform the procedures on the clips mentioned in **Section 6 1** are tabulated below

Clip No (channel)	Programme content description	Length of clip	Time taken to complete analysis (percent of clip length)
1 (A)	Chat Show	600 seconds	40 seconds (6 6%)
2 (A)	News Broadcast	1200 seconds	83 seconds (6 9%)
3 (A)	Music Show	600 seconds	35 seconds (5 8%)
4 (B)	News Broadcast	1200 seconds	77 seconds (6 4%)
5 (B)	Soap Opera	1200 seconds	73 seconds (6 0%)
6 (B)	Sports Show	1200 seconds	83 seconds (6 9%)
7 (B)	Cookery Show	2400 seconds	167 seconds (6 9%)
8 (C)	Sports Show	1800 seconds	124 seconds (6 8%)
9 (C)	Youth Magazine	1800 seconds	119 seconds (6 6%)
10 (C)	Game Show	1500 seconds	82 seconds (5 4%)
11 (C)	Comedy Quiz	1800 seconds	105 seconds (5 8%)
12 (D)	News Broadcast	1800 seconds	102 seconds (5 6%)
13 (D)	Cartoon	1800 seconds	121 seconds (6 7%)
14 (D)	Music Show	1800 seconds	106 seconds (5 8%)
15 (D)	Nature Show	2400 seconds	154 seconds (6 4%)
16 (A)	News Broadcast	1200 seconds	69 seconds (5 7%)
17 (B)	Sci-fi Show	2700 seconds	184 seconds (6 8%)
AVERAGE	-------	-------	6 4%

APPENDIX-C: PUBLICATIONS & APPENDED PAPERS

AUTOMATIC TV ADVERTISEMENT DETECTION FROM MPEG BITSTREAM
Sadlier, D , Marlow, S , O'Connor, N , Murphy, N Workshop on Pattern Recognition in Information Systems 2001 (in conjunction with the Third International Conference on Enterprise and Information Systems – ICEIS 2001) Setubal, Portugal, July 2001

AUDIO AND VIDEO PROCESSING FOR AUTOMATIC TV ADVERTISEMENT DETECTION
Marlow, S , Sadlier, D , McGeough, K , O'Connor, N , Murphy, N Irish Signals & Systems Conference, National University of Ireland, Maynooth, June 2001

Automatic TV Advertisement Detection from MPEG Bitstream

David A Sadlier, Dr Sean Marlow, Dr Noel O'Connor and Dr Noel Murphy

Centre for Digital Video Processing/Research Institute for Network & Communication Eng
Dublin City University
Rep of Ireland
sadlierd@eeng dcu ie

Abstract. The Centre for Digital Video Processing at Dublin City University conducts concentrated 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 *Fischlar* [1], which provides for efficient recording, analysing, browsing and viewing of digitally captured television programmes Advertisement breaks during or between television programmes are typically recognised by a series of 'black' video frames simultaneously accompanying a depression in audio volume which separate each advertisement from one another by recurrently occurring before and after each individual advertisement It is the regular prevalence of these flags that enables automatic differentiation between what is programme and what is a commercial break This paper reports on the progress made in the development of this idea into an advertisement detector system that automatically detects the commercial breaks from the bitstream of digitally captured television broadcasts

1 Introduction

For the development of an efficient video browsing/viewing tool for digitised television, it is desirable to present the user with the option of skipping irrelevant content A typical television programme may be accompanied by beginning/end credits with one or more ad-breaks somewhere in the middle To the user, these features of a programme/video are generally regarded as an insignificant part of the recorded material Hence, by detecting the ad-breaks, the efficiency of programme browsing/viewing may be increased

Advertisement breaks may be isolated from actual programme material by the flags that most terrestrial and some satellite television companies wave during their broadcast a series of 'black' video frames simultaneously accompanied by a decrease in the audio signal occurring before and after each individual advertisement [2]

The *Fischlar* 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 platform

An inherently dark or 'black' frame of a video may be recognised by its luminance histogram, which would be typically characterised by having most of its 'power' at the bottom end of the pixel amplitude spectrum, corresponding to black/dark pixels

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, an audio depression (of magnitude defined by threshold) may be detected
The abovementioned criteria propose a simplistic approach to the task of locating within a television programme, groups of black-frames/audio-depressions, which should provide for efficient 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 [MPEG-1 (Layer-II)], which is highly undesirable from a computational point of view [2]
It was proposed 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 (DC-DCT) 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' [The Discrete Cosine Transform (DCT) forms an integral part of the compression process used in the MPEG-1 standard]

- For Audio, an inspection of the weight of the scalefactors of the signal's (low) frequency subbands with a view to establishing whether or not a video frame's accompanying audio signal power is minimal enough for it to be labeled 'silent' [In the MPEG-1 Layer-II standard for compression of audio signals, the signal frequency spectrum is divided uniformly into 32 subbands which are then coded independently from one another]

It was envisaged that black-frame/silence detection via the abovementioned principles would provide for advertisement detection from captured and encoded television programmes to the same degree of accuracy as could be obtained by the methods requiring implementation of a full audio/visual decode, but with the computational burden significantly reduced

2 Background

The Digital Video Indexing Project at Dublin City University is a continuing research endeavor aimed at developing innovative technologies fundamental to the realisation of efficient video content management

To demonstrate our research on digital video, we have developed a Web-based digital video system which we call *Fischlar* [from the Irish *fis* (vision) and *chlar* (programme)] 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 Video Cassette Recorder
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
[SMS Short Messaging Service, WAP Wireless Application Protocol, PDA Personal Digital Assistant]

3 'Black/Silent' Frames

3 1 The MPEG Standard

The Moving Pictures Experts Group (MPEG), who meet under the International Standards Organisation (ISO), generate the international standards for digital video and audio compression MPEG-1 [3] is a standard in five parts which individually address the issues of audio/visual multiplexing, video coding, audio coding, bitstream testing, and software implementation A detailed description of the MPEG-1 standard for audio and video compression may be found via references [3] & [4] and is thus not dealt with here

3 2 Black Frame Detection

Fischlar captures analog television signals and encodes them according to the MPEG-1 standard An MPEG-1 video frame is divided into slices, which are subdivided into macroblocks which each contain 6 blocks of (8x8) pixels transformed by a 2D-DCT, 4 of which provide luminance information, leaving 2 for chrominance information
The video data is either explicitly given for the frame (I-frame) or provided implicitly via forward prediction (P- frame) or forward/backward prediction (B-frame)
The four luminance blocks (Y-blocks) of each macroblock provide the essential information on how dark or 'black' the macroblock effectively is, hence indicating how dark or 'black' the overall frame may be
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, to which the visual perception is highly sensitive, whereas, various frequency fluctuations may go unnoticed 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, i e AC variations may be ignored
The proposal was that an average luminance intensity value for each frame could be determined from the DC-DCT coefficients provided within individual Y-block This value was expected to be relatively low for inherently dark frames and higher for brighter frames Thus by thresholding this value, a decision on whether the frame is 'black' or not may be made

3 3 Silent Frame Detection

Fischlar captures television audio signals and encodes them according to the MPEG-1 Layer-II compression algorithm which encodes as follows
The frequency spectrum of the audio signal (sampled at 32, 44 1, or 48kHz) is first divided uniformly into 32 subbands which approximate the ear's critical bands These subbands are then individually assigned a bit-allocation according to the audibility of quantisation noise within that band (a pyschoacoustic model of the ear analyses the audio signal and provides this information to the quantiser)
Layer-II frames consist of 1152 samples, 3 groups of 12 samples from each of 32 subbands A group of 12 samples gets a bit allocation and, if this is non-zero, a scalefactor Scalefactors are weights that scale groups of 12 samples such that they fully use the range of the quantiser (the encoder uses a different scalefactor for each of the three groups of 12 samples within each subband only if necessary) 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 it provides an indication of the maximum power exhibited by any one of the 12 samples within the group
The proposal was that an audio power level for each video frame could be determined by superposition of the scalefactors corresponding to the groups of audio samples to which the frame is associated This power level was expected to be relatively low for 'silent' frames and higher for louder frames Thus by thresholding this value, a decision on whether the frame is 'silent' or not may be made
Further still, it was expected that examination of just the bottom end of the frequency spectrum would provide sufficient information on which this decision could still be accurately made, since it is typical of an audio signal to have most of its energy corresponding to relatively low frequencies
By trial and error examination of various audio signals, it was decided that at least 10 of the low frequency subbands' scalefactors must be included in the silence investigations such that the results rendered were both sensible and consistent
[Since the maximum frequency encoded by the MPEG-1 Layer-II standard is 20kHz, the first 10 subbands corresponds to an examination of the energy present from 0-6kHz]

4 Ad-break Detection

4 1 Recognition of Advertisement breaks

As explained, the occurrence of black-frame/audio-depression series may indicate the existence of an ad-break However, it is possible, and maybe quite probable, that these indicators also occur during the valuable material of 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 'black/silent' frame series occurring between individual advertisements tended to be of at least 6 frames in length Thus to aid against detection of freak black-frame/audio-depression occurrences not associated to ad-breaks which may sporadically occur during programmes, it was decided only to recognise them if they exhibit a series of **at least 6 consecutive 'black/quiet' frames** i e a series of 10 'black/quiet' frames between adjacent advertisements would be detected, while a series of 5 occurring between anchor person and news report would not

- Upon examination of 20+ advertisement breaks from various television stations, the longest advertisement recorded was that of 76secs, with approximate average advertisement duration of 25secs Thus it was decided that upon detection of a series of (at least 6) 'black/silent' 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 This prevented against recognition of rogue 'black/quiet' frame series which may randomly occur during relevant programme material
  A consequence of this condition was that the system would fail upon occurrence of advertisements that were longer than 90secs in duration (76secs was longest advertisement recorded so 90secs provided for some tolerance)
  Frame rate for *Fischlar* is 25 frames/sec Thus 90secs corresponds to 2250 frames

- Upon examination of 20+ advertisement breaks from various television stations, it was determined that the average number of individual advertisements within an ad-break was approximately seven
  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**
  i e upon detection of a series of (at least 6) 'black/quiet' frames, at least three more series must be detected, within the respective 90sec windows of each other, for the overall detection to be recognised as an ad-break (detection of 4 'black/quiet' frame series corresponds to 3 individual

advertisements) This prevented against possible mistaken recognition due to the unlikely event of up to three rogue series (consisting of at least 6 consecutive 'black/quiet' frames) occurring within 90secs of each other, during relevant programme material

The above three conditions would be individually weak since they focus on features which are undoubtedly inconsistent attributes of a television broadcast However it is the combined effect of all three clauses which is expected to provide the success in accurately preventing mis-recognition of programme content for advertisement material Figure-1 explains how detection of sporadic 'black/quiet' frame series are interpreted and recognised (shaded frames represent frames which are both 'black' and 'silent')

4 2 Black-frame & Silent-frame Thresholds

A number of threshold techniques were investigated However, the following adaptive method provided the most consistency in the results obtained, and was thus chosen as the appropriate scheme

For video, an overall mean DC-DCT value was calculated by averaging over all individual frames (= $DC\text{-}DCT_{avg}$) The black-frame' threshold was then expressed as some factor times this number By trial and error examination of various ad-break clips, the video threshold which gave the most sensible and consistent results was

$$V_{th} = 0\ 48 * DC\text{-}DCT_{avg} \quad (1)$$

For audio, an overall mean audio level value was calculated by averaging over all individual frames (= $audio\_level_{avg}$) The 'silent-frame' threshold was then expressed as some percentage of this number By trial and error examination of various ad-break clips, the audio threshold which gave the most sensible and consistent results was

$$A_{th} = 0\ 073 * audio\_level_{avg} \quad (2)$$

4 3 Procedure

4 3 1 Video Examination

- The DC-DCT coefficients of each Y-block of each frame were stripped from the video bitstream of the MPEG file
- An average luminance DC-DCT coefficient was then calculated for each video frame of the sequence
- The overall mean value of the average frame coefficients for the clip was determined The 'black-frame' threshold was then defined as the value corresponding to 48% of this number [see (1)]
- Each frame's average DC-DCT coefficient value was compared to the threshold and if equal/less than, then the frame was labeled 'black'

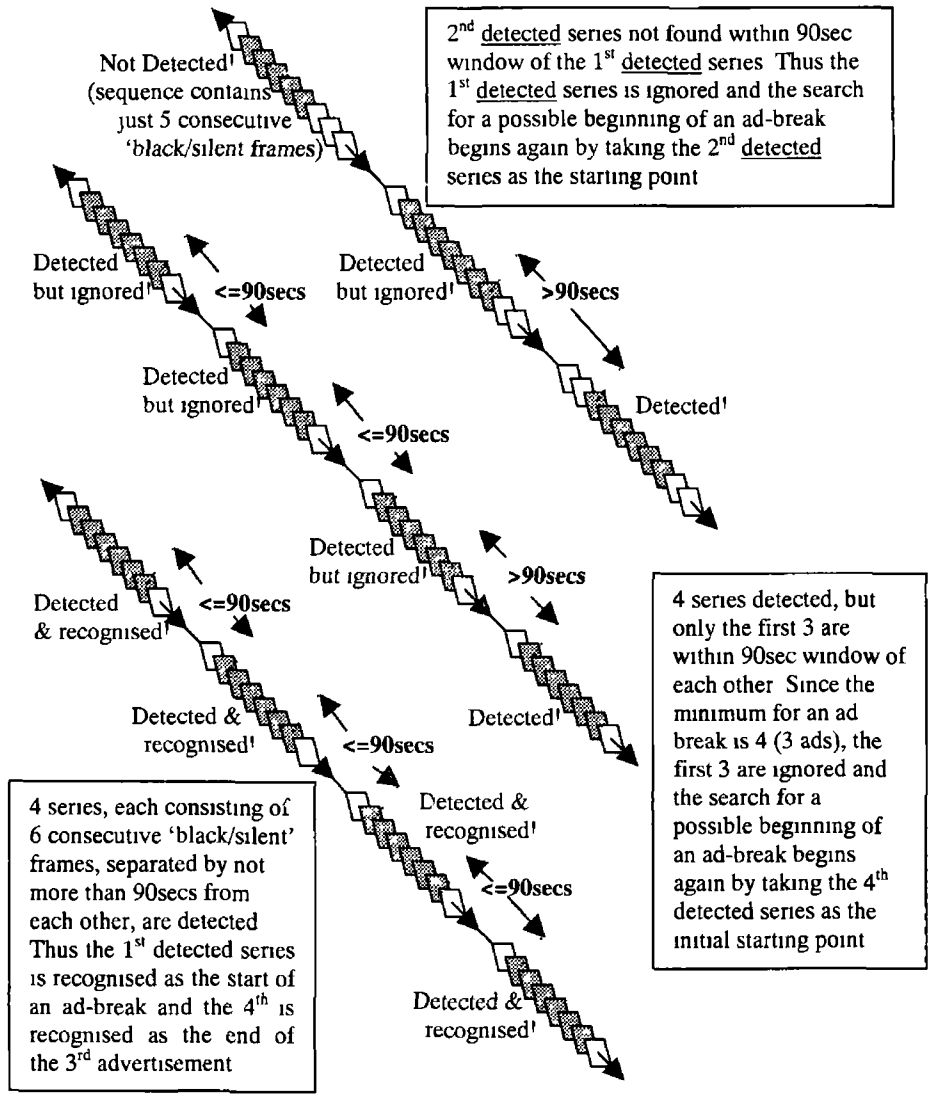


Fig 1 Interpretation of 'black/silent' frame series

4 3 2 Audio Examination

- The scalefactors corresponding to the first 10 subbands of the encoded audio signal were stripped from the bitstream
- An audio level for each video frame was determined by averaging the scalefactors corresponding to its associated audio signal
- The overall mean value of the video frame audio levels for the entire clip was determined The 'silent-frame' threshold was then defined as the value corresponding to 7 3% of this number [see (2)]

- Each video frame's representative audio level was compared to the threshold and if equal/less than, then the frame was labeled 'silent'

4 3 3 Search for Simultaneous 'Black/Silent' Frames

- Unless 4 distinct series of at least 6 consecutive simultaneously 'black/silent' frames were detected within 90secs (2250 frames) of each other, any such series were ignored
- Upon detection of 4 distinct series of at least 6 consecutive simultaneously 'black/silent' frames, within 90secs (2250 frames) of each other, the minimum requirements for such a sequence to be recognised as an ad-break have been completed
- Further such series detected within the same allowable window [90secs (2250 frames)] from the end of the previous would also be recognised as being part of the same ad-break
- Until there is no further detection of such series within the allowable window, all detected series are recognised as corresponding to the same ad-break
- The detected ad-break is then said to have begun with the first 'black/silent' frame of the first detected/recognised series and ended with the last frame of the last detected/recognised series
- The process begins again upon detection of the next (unrelated) series

5 Results & Examination

5 1 Test Material

Fischlar provided 10 short television programme clips from 4 different channels [labeled (a), (b), (c) & (d)] in MPEG-1 Layer-II format The recordings were meticulously chosen such that they exhibited significant content diversity with at least one complete ad-break somewhere in the middle The abovementioned procedures were executed on all 10 clips Evaluation of the system is performed by comparison of results achieved against a manual record of the true location of the ad-breaks, to the nearest second, within each clip

5 2 Results

For shorthand purposes, a detected series of at least 6 consecutive 'black/silent' video frames as previously described, is henceforth labeled as a **flag**

(channel) Description	Length of clip (seconds)	# Flags detected	# Flags detected, not corresponding to ad-breaks	True ad-break location in clip (second second)	Detected ad-break location in clip (second-second)
(a) Chat Show	600s	5	0	193s – 281s	193s – 281s
(a) News Broadcast	1200s	4	0	468s – 538s	468s – 516s
(a) Music Show	600s	5	1	257s – 347s	258s – 347s
(b) News Broadcast	1200s	11	0	113s – 351s	113s – 351s
(b) Soap Opera	1200s	7	0	779s – 935s	779s – 935s
(b) Sports Show	1200s	11	3	603s – 790s	603s – 770s
(c) Youth Magazine	1800s	14	7	185s – 272s	185s – 272s
(c) Game Show	1500s	7	2	659s – 772s	659s – 772s
(d) Cartoon	1800s	21	1	798s – 975s and 1415s – 1660s	798s – 975s and 1415s – 1660s
(d) Comedy Quiz	1800s	6	0	774s – 934s	774s – 913s

Table 1 Results of experimentation on 10 video clips provided by 4 television stations comprising 11 advertisement breaks

5 3 Result Examination

The absence of **falsely identified** content is evident from the results in Table-1
In clip '(b) Sports Show', the detected end of the ad-break is at 770 seconds, but the true end is later, at 790 seconds This is recorded as 20 **missed** seconds of ad-break material

The total number of seconds in the true ad-break (= 790 – 603) = 187 seconds
The total number of seconds detected as ad-break (= 770 – 603) = 167 seconds
A superposition of these quantities over all 10 clips was performed to give four overall values, which are illustrated below

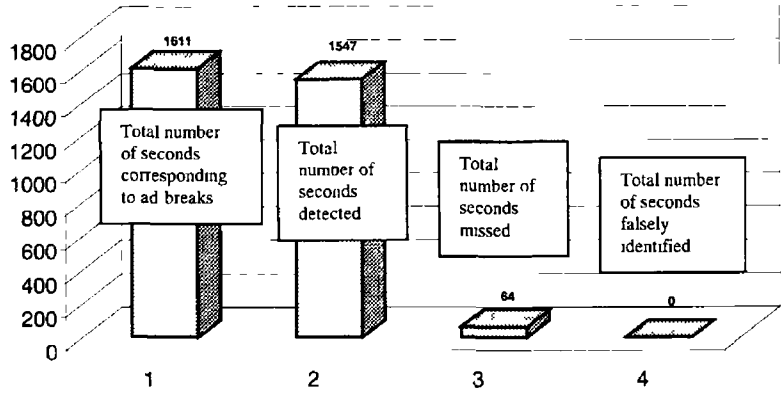


Fig 2 Total number of seconds corresponding to ad-breaks, detected ad-breaks, missed ad-breaks and falsely identified ad-breaks

5 3 1 Precision & Recall

For a better insight into the individual accuracy of each clip, two important figures of merit for the ad-detection system were calculated

- The **Recall** measure looks at the percentage of detected material corresponding to true ad-breaks

$$\textbf{Recall} = \frac{\textbf{100 * [Length of ad-break (seconds) – No seconds missed]}}{\textbf{Length of ad-break (seconds)}}$$

- The **Precision** measure is a percentage showing how accurate the system is at exclusively detecting ad-break material

$$\textbf{Precision} = \frac{\textbf{100 * [Length of ad-break (seconds) – No seconds missed]}}{\textbf{Length of ad-break (seconds) – No seconds missed + No seconds falsely identified}}$$

Example

For clip '**(b)** Sports Show the Recall/Precision figures were calculated as follows (from the information in Table-1)

Length of ad-break (= 790 – 603) = 187 seconds

No seconds falsely identified = 0 seconds

No seconds missed = 20 seconds (ad-break-end detected 20 seconds early)

Precision = 100 * [(187 – 20) / (187 – 20 + 0)] = **100**

Recall = 100 * [(187 – 20) / 187] = **89 3**

Results following similar calculations performed on all clips are presented in Table-2

Clip	Precision	Recall
(a) Chat Show	100	100
(a) News Broadcast	100	68 6
(a) Music Show	100	98 8
(b) News Broadcast	100	100
(b) Soap Opera	100	100
(b) Sports Show	100	89 3
(c) Youth Magazine	100	100
(c) Game Show	100	100
(d) Cartoon (1)	100	100
(2)	100	100
(d) Comedy Quiz	100	86 9

Table 2 Precision & Recall values based on results in Table 1

6 Conclusions

6 1 Result Evaluation

In all, 10 clips comprising 315 minutes of digital video, incorporating 11 ad-breaks, were analysed The following are the main points to be noted

- The system detected the occurrence of all 11 ad-breaks
- A total of 14 'black/silent' flags which were not associated with ad-breaks were detected (7 of these occurring in one clip alone) However, the number of falsely detected ad-breaks was zero
- All detections attained a Precision percentage of 100%
- 8-out-of-11 of the detections attained a Recall percentage greater then 98%

Clips '(b) Sports Show, (d) Comedy Quiz, & (a) News Broadcast' performed relatively poorly compared to others It was noted that their common downfall was the ad-break-end being detected prematurely Manual investigation showed that the reason for this was that for all three clips, the final flag in the ad-breaks (occurring between final ad and return of programme) consisted of less than 6 consecutive frames, thus violating one of the three conditions imposed on the detection process Consequently, the system did not detect the occurrence of the final advertisement within the ad-break, but identified the ad-break-end with the end of the last detected flag (which actually corresponded to the end of the penultimate advertisement)
Numbers of individual advertisements comprising overall ad-breaks were counted for the three clips It was then concluded that clips '(b) Sports Show' and '(d) Comedy Quiz' resulted in 1-out-of-7 and 1-out-of-6 ads missed within their respective ad-breaks, giving Recall percentages of **89 3** and **86 9** However, clip '(a) News Broadcast' resulted in 1-out-of-4 ads missed within its ad-break, which represented non-detection of a much larger proportion of the overall ad-break Hence, this clip only attained a meagre Recall percentage of **68 6**
The consistently high Precision figures indicated high success in the prevention of mis-recognition of content not corresponding to ad-breaks

6 2 Further Work

The system succeeded very well until, on 3-out-of-11 occasions, the conditions incorporated to combat false identification of ad-breaks, actually prevented the detection of the final advertisement within their respective ad-breaks For this reason (and due to the fact that the system will fail for (i) occurrence of individual advertisements longer than 90secs in duration, and (ii) ad-breaks consisting of less than 3 advertisements), any future work could possibly involve further optimisation of the trade-off between consistently precise ad-break detection and the prevention of false identification, either by changing the parameters in the existing conditions or by replacing them with improved ones

In introducing this subject, it was mentioned that most advertising television stations feature the characteristic of 'black/silent' frame gap in between individual advertisements In fact, of the six advertising television channels captured by *Fischlar*, two do not exhibit this trait They instead have each individual advertisement run directly into each other with no pauses in between Consequently, the discussed method of ad-break detection would fail for these broadcasts and so an alternate method is required

A proposed technique is to examine the rate of shot-cuts over an entire programme This rate is expected to be notably high during ad-breaks since advertisements characteristically exhibit a consistently high rate of activity This, coupled with some timing aspects (e g broadcast time is usually sold in discrete units) could provide sufficient information enabling an alternative method for automatic advertisement/programme differentiation

References

1 *Centre For Digital Video Processing/Fischlar Website* http //www fischlar com

2 Carroll, D *Advertisement Detection* Internal technical report, Dublin City University 2000 (contact sadlierd@eeng dcu ie)

3 *The Official MPEG Committee Website* http //www cselt it/mpeg/

4 *Video and Audio Compression Website* Dept of Computer Science, Cardiff University Website http //www cs cf ac uk/Dave/Multimedia/node196 html

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

6 NOMAD Website http //www fatalfx com/nomad/

7 Lienhart, R , Kuhmunch, C & Effelsberg, W University of Mannheim *Detection & Recognition of Television Commercials* http //www informatik uni-mannheim de/informatik/pi4/publications/library/Lienhart1996e pdf

8 Fry, D , Hampshire, E , Hargrove, T *Automatic Detection of Commercials within Pre-Recorded Cartoon Shows* http //www cse scu edu/projects/2000-01/project12/

9 Li, Y & Jay Kuo, C -C *Detecting the Commercial Breaks in Real TV Programs Based on Audiovisual Information* http //viola usc edu/newextra/Research/Ying_Maverick/Ying_publications htm

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 *Fischlar* [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 *Fischlar* [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 Fischlar) 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 *Fischlar* 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

Fischlar 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

Fischlar 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

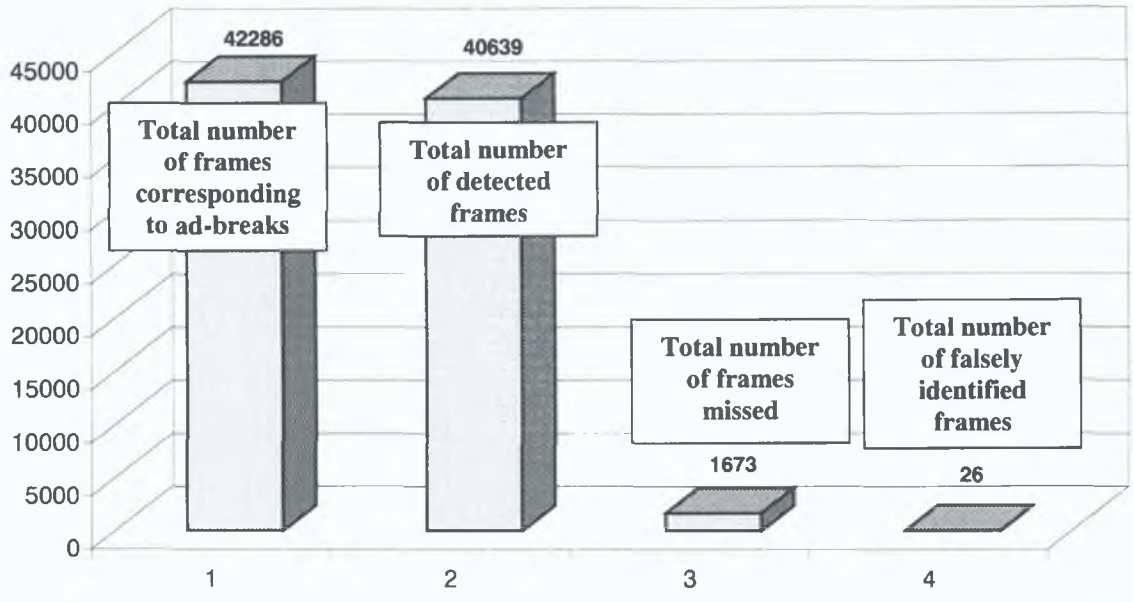


Figure-1 Total number of frames corresponding to ad-breaks, detected ad-breaks, missed frames and falsely identified frames.

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 Fischlar, 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

Fischlar 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/Fischlar Website* http //lorca compapp dcu ie/Video/

2 Lee H, Smeaton A, O'Toole C, Murphy N, Marlow S and O'Connor N *The Fischlár 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