

3. COMMERCIAL-CLASSIFICATION PROCESS

In a first step the visual data is extracted from the mpeg-2 streams.

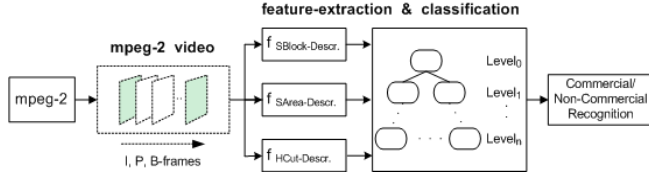


Figure 3: Process involved computing the feature vectors

We consider all consecutive frames of the image sequence and save them temporarily for processing.

3.1. Visual-Descriptors

German commercials have several characteristic features in common. They are shown in Figure 4, like:

- (a) appearance of 3 to 14 separating frames between each commercial block,
- (b) duration of individual commercials more than 10 and less than 60 seconds,
- (c) removal of the network-logo during the commercial breaks and
- (d) high number of cuts within the spots.

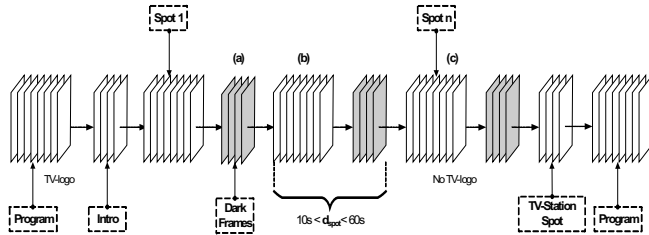


Figure 4: Structure of a commercial block

Based on these observations, we designed three descriptors to transform our data into feature vectors.

3.1.1. Separating-Block Descriptor

The descriptor is depicted in Figure 5. We determine the average brightness L_μ as well as the variance L_{var} of the pixels for each frame. All candidates as separating frames have to be below a certain threshold. Next we examine the average brightness $L_{\mu-sb}$ for 3x3 sub-images of the selected frame. The number of consecutive frames, satisfying these requirements, is counted and has to be in a certain range (a). This is called a separating block f_1 , if the time distance to the previous separating block fulfills the restrictions (b) regarding the duration of individual commercials.

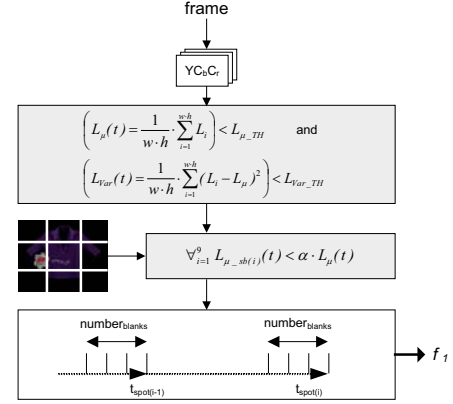


Figure 5: Block diagram of the separating-block descriptor

3.1.2. Static-Area Descriptor

The descriptor is depicted in Fig. 6. We scan the areas c_0-c_3 , where a TV-logo is displayed usually. The corresponding brightness values of the first frame of a processing window N are saved in L_{buffer} . For each consecutive frame, we compare the current values $L_{actFrame}$ with L_{buffer} and save the darkest values of the search areas. After a length of N frames, a binary image is generated by comparing the resulting values to a threshold. If the average brightness value μ_c of at least one of the four corners is higher than zero, we detected static pixels f_2 and the probability of a logo, present in this area, is high.

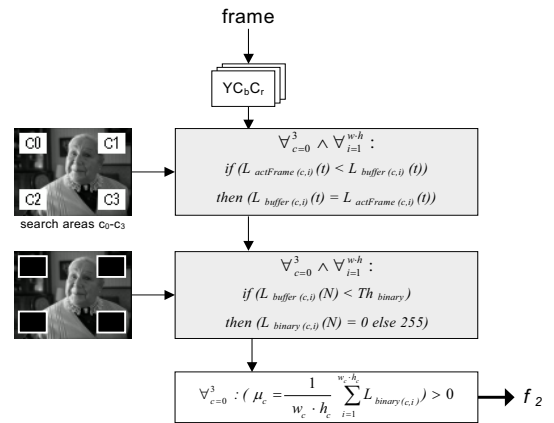


Figure 6: Block diagram of the static-area descriptor

3.1.3. Hard-Cut Descriptor

The difference of two consecutive I-frames presents a hard cut, if this difference exceeds a certain threshold. The amount of these values is averaged over a window N resulting in the hard cut frequency f_3 (Fig.7). We implemented also one descriptor which builds on the motion-activity information included in the mpeg-2 stream.

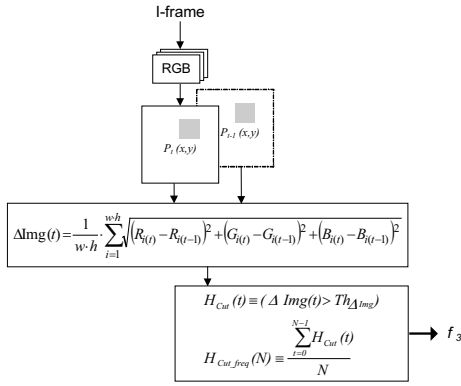


Figure 7: Block diagram of the hard-cut descriptor

We applied the descriptors to the training-data. The extracted features of the commercial sequences appeared within a decision window of 50 I-frames mainly in three combinations, which is shown in Figure 8.

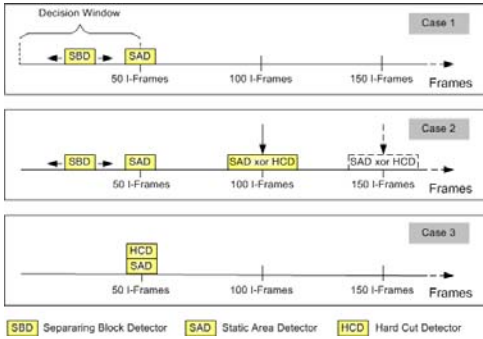


Figure 8: Appearance of features in commercials

In case1 a separating-block and non-static areas (no-logos) were detected within a decision window.

In case2 either a high hard-cut frequency or no-logo was detected providing that case1 occurred within the last 100 I-frames. In the third case no-logo and a high hard-cut-frequency were simultaneously detected.

3.2. Decision-Tree

These cases are combined with a decision tree according to the ID3 algorithm [10]. We calculate the average entropy E for each case from the training-data shown in Table 1.

$$E_{average} = \sum_{j=1}^b \left(\frac{n_b}{n_t} \right) \cdot \left[\sum_{i=1}^c - \left(\frac{n_{bc}}{n_b} \right) \log_2 \left(\frac{n_{bc}}{n_b} \right) \right]$$

	Commercial			Non-Commercial		
	correct accept.	false accept.	Σ	false reject.	correct accept.	Σ
case 1	82	0	82	16	384	400
case 2	8	0	8	90	348	474
case 3	56	2	58	42	382	424
	n_{bc1+}	n_{bc2+}	n_{b+}	n_{bc1-}	n_{bc2-}	n_{b-}

Table 1: Cases with accepted/rejected rate as commercial

The case with the lowest entropy forms the first node. The resulting decision tree is shown in Figure 9.

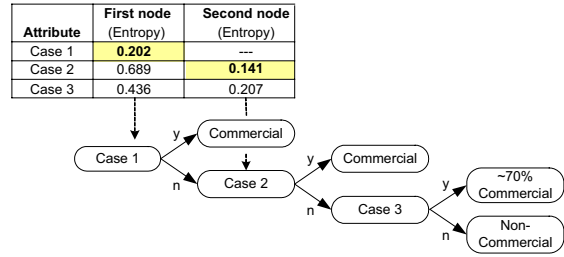


Figure 9: Block diagram of the developed decision tree

4. EXPERIMENTS

The experiments were carried out on a database of 200 representative video sequences (100 sequences for training and 100 as testing-data), in total of 400 min of recordings; 40 ‘commercials’ and 4*40 ‘non-commercials’ (cartoon, music, news and sport) of 2 minutes’ each gathered from popular TV in Germany (ARD, ZDF, BBC, VIVA..). We extracted frames and scaled them down to a resolution of 90*72 pixels. The number of considered frames in the processing window is $N=50$ I-frames.

4.1. Experimental Results

We examined the performance of each descriptor on the training-data. The separating-block descriptor should detect the transition from a running spot to a new spot.

Genre	observed number of sep. blocks	detected number of sep. blocks
Commercial	111	108
Cartoon	0	2
Music	0	7
News	0	0
Sport	0	0

Table 2: Classification accuracy of the sep.-block descriptor

From 111 subjectively determined separating blocks within the 20 ‘commercials’ only three blocks, including fades, were misclassified. The same reason caused the false acceptances in ‘cartoon’ and ‘music’. In ‘news’ and ‘sports’ the descriptor worked correctly.

Table 3 shows the amount of windows N within each genre and the detected number of windows with non-static areas.

Genre	examined number of windows	detected number of non -static areas/ no logo
Commercial	98	90
Cartoon	94	0
Music	105	1
News	95	1
Sports	90	0

Table 3: Classification accuracy of the static-area descriptor

From 98 windows in commercial, 90 with non-static-areas were detected. The remaining 8 contained a static ‘company-logo’. In ‘music’ we had a window including a scene-change without the TV-logo and in ‘news’ the logo was at the beginning on a window outside the scanned corners.

Table 4 shows the performance of the hard cut descriptor.

Genre	examined number of windows	detected num. of windows with high-cut rate
Commercial	98	59
Cartoon	94	20
Music	105	45
News	95	4
Sports	90	15

Table 4: Classification accuracy of the hard-cut descriptor

The recognition of ‘commercials’ in a processing window using a single descriptor is obviously not sufficient. In order to achieve high identification rates, we developed the tree with nodes based on logical combinations of the descriptors and a branching ratio of 2.

Tab. 5 shows the detection rates for the 20 video sequences of each genre used in our experiments.

Genre	examined number of windows	detected number of commercials	(%)
Commercial	98	91	93%
Cartoon	94	0	0%
Music	105	1	1%
News	95	1	1%
Sports	90	0	0%

Table 5: Probability for video being detected as Commercial

In our experiment 91 from 98 examined windows from the genre ‘commercial’ were classified as commercial.

The remaining 7 windows were very close to the commercial-detection threshold of 50%. Those sequences started within a commercial, after the appearance of a separating block, presented a long spot with a ‘company-logo’ and had a low high-cut-frequency. It is obvious, that detection of sequences with ‘special cases’ is highly unreliable.

It is interesting to note that only 1 from 105 examined windows of the genre ‘music’ and 1 of 95 windows of ‘news’ were misclassified (caused by non-static areas and high cut frequency). The genre ‘cartoon’ and ‘sport’ were correctly classified in more than 90 windows.

5. SUMMARY & CONCLUSION

We presented a new approach for the detection of commercials. Three key contributions have been made. We started with new or modified visual descriptors. Second, we considered the temporal relations of the features. Third, we developed a decision tree to combine the results of the visual descriptors, deriving a probability rate for a video sequence being a ‘commercial’ or ‘non-commercial’.

A video database containing five popular genres namely cartoon, commercial, music, news and sports has been used. An average correct classification rate of 93% for commercial-videos detected as a ‘commercial’ and more than 99% for the other genres detected as a ‘non-commercial’ has been achieved.

With our current software system we achieved on an AMD Athlon XP1600+, 1.41 GHz a run-time performance of approximately 1 min for classification for 1 min of video.

11. REFERENCES

- [1] A. Albiol, M.J.C. Fulla, F.A. Albiol and L. Torres, “Detection of TV commercials”, IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 3, pp. 541 - 544, 2004.
- [2] B. Satterwhite and O. Marques, “Automatic detection of TV commercials”, IEEE Potentials, pp. 9 – 12, 2004.
- [3] J.H. Yeh, J.C. Chen, J.H. Kuo and J.L. Wu, “TV Commercial Detection in News Program Videos”, IEEE International Symposium on Circuits and Systems, pp. 4594 - 4597, 2004.
- [4] R.J.M. den Hollander and A. Hanjalic, “Logo recognition in video stills by string matching”, International Conference on Image Processing, volume 3, pp. 517 - 20, 2003.
- [5] D. A. Sadlier, S. Marlow, N. O’Connor and N. Murphy, “Automatic TV Advertisement Detection from MPEG Bitstream”, International Conference on Enterprise Information Systems, pp. 449 - 452, 2001.
- [6] R. Lienhart, C. Kuhmüch and W. Effelsberg, “On the Detection and Recognition of Television Commercials”, IEEE Conference on Multimedia Computing and Systems, pp. 509 - 516, 1997.
- [7] R. Glasberg, K. Elazouzi and T. Sikora, “Video-genre-classification: recognizing cartoons in real-time using visual-descriptors and a multilayer-perceptron”, The 7th International Conference on Advanced Communication Technology, volume 2, pp. 1121 - 1124, 2005.
- [8] M. Roach, J. Mason and L.-Q. Xu, “Video genre verification using both acoustic and visual modes”, International Workshop on Multimedia Signal Processing, pp. 157 – 160, 2002.
- [9] B. T. Truong, S. Venkatesh and C. Dorai, “Automatic genre identification for content-based video categorization”, Proceedings 15th International Conference on Pattern Recognition, volume 4, pp. 230 - 233, 2000.
- [10] J.R. Quinlan, “Decision trees and decision-making”, IEEE Transactions on Systems, Man and Cybernetics, volume 20, pp. 339 - 346, 1990.
- [11] T. Sikora, P. Salembier and B.S. Manjunath, Introduction to MPEG-7: Multimedia Content Description Interface, John Wiley LTD, ISBN 0471486787, 2002.