Real-Time Approaches for Video-Genre-Classification using New High-Level Descriptors and a Set of Classifiers

Ronald Glasberg, Sebastian Schmiedeke, Martin Mocigemba and Thomas Sikora
Communication Systems Group
Technical University Berlin
10587 Berlin, Germany
{glasberg, schmiedeke, mocigemba, sikora}@nue.tu-berlin.de

Abstract

In this paper we describe in detail the recent publications related to video-genre-classification and present our improved approaches for classifying video sequences in real-time as 'cartoon', 'commercial', 'music', 'news' or 'sport' by analyzing the content with high-level audio-visual descriptors and classification methods. Such applications have also been discussed in the context of MPEG-7 [1]. The results demonstrate identification rates of more than 90% based on a large representative collection of 100 videos gathered from free digital TV and Internet.

1. Introduction

The great challenge in the field of multimedia content analysis is the transformation of human interpretations of audio-visual data to the respective machine processable representation. The difference between these two spheres is the so called 'semantic gap'. Bridging this gap will open up a wide field of new applications. One possible application is the content selection in TV and world wide web according to user-specific profiles, e.g. genres like cartoon, commercial, music, news and sport. Humans perceive genres as patterns of audio-visual sequences describing dimensions like narration, aesthetics etc. Our research focuses therefore on descriptors and classification methods used for describing these patterns according to Fig. 1.

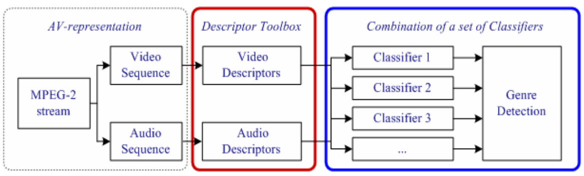

Figure 1. Flowchart of descriptors and classifiers

2. Genre Analysis

In the European Union 98% of all households regularly watch TV or enjoy videos in the *world wide web*. Among others, popular genres are cartoon, commercial, music video, news and sport. In order to reliably detect these genres without manual indexing, research has focused on two different approaches, two-category $\Omega_2=\{c, \underline{c}\}$ and multi-category $\Omega_m=\{c_1, \ldots, c_m\}$ detection by applying various descriptors and classification methods like neural networks, bayes, decision trees, hidden markov models and support vector machines.

3. Cartoons

The characteristics of cartoons primarily originate from the method of production. Specific techniques using drawn pictures or computer generated images determine the aesthetics and enable humans to differentiate between natural images and cartoons.

3.1. Related Work

To detect cartoons, [6] introduce a qualitative cartoon/non-cartoon spectrum ranging from videos with 'strong' cartoon aesthetics to videos without these characteristics (real life) according to Table 1.

Table 1. Publications related to cartoon detection

Publ.	Genres	Database & Cliplength	Video-features	Audio-features	Classifier	Results ($T_{Decision}$)
[3]	CAR, <u>CAR</u>	5x 20 Clips à 2 min	Brightness, Color Nuance, Saturation, Edge, Motion Activity	MFCC	GMM (audio), MLP (video)	Recall = 90% on Clip-length
[4]	CAR, <u>CAR</u>	13 Clips à 74 min	Dominant Color Ratios (light/dark etc.), Global Weight Color Histogram	/	/	Pure Descriptor Performance
[5]	CAR, PHOTO	14.000 jpeg (1.620 CAR)	Brightness, Color Histogram, Saturation, Edge, Compression Ratio, Granulometry	/	SVM	CA = 94%
[6]	CAR, <u>CAR</u>	8x CAR, 20x NCAR à 40 s	Motion Activity and Acceleration	/	GMM	CA=90% (5s) CA=96% (25s)
our result	CAR, <u>CAR</u>	5x 20 Clips à 3 min	Logo (Detection, Recognition), Dominant Color, Nuance, Color Count	/	HMM	CA=96% (20 s)

The aesthetic of cartoons is characterized by strong colors, patches of uniform color, strong black edges and motion. Based on these features the following descriptors have been developed (Table 1) to transform the audio-visual data into feature vectors:

Brightness-Descriptor: In [3] a descriptor-family is presented, extracting the average brightness f_{BAvg}, the percentage of pixels with brightness values exceeding a threshold f_{BTh} and brightness changes f_{BFreq} of consecutive frames. [5] analyze f_{BTh} for captured frames and [30] in the context of multi-category detection for consecutive frames.

Color Saturation-Descriptor: According to the descriptor-family mentioned above, a further family was implemented in [3], extracting the average saturation values f_{SAvg} [5], threshold f_{STh} [30] and change f_{SFreq} of consecutive frames.

Color Distribution-Descriptor: [5] exploit the appearance of few colors in cartoons to define a descriptor employing color frequencies from histograms. [30] consider the same descriptor in the multi-category case.

Patches of uniform Color-Descriptor: Objects in 'traditional' cartoons are perceived as regions of same color. Two methods analyze this feature:

a) Color Nuance Descriptor [3]: This descriptor determines the mean color distance between a pixel and its adjacent 8 pixels. The average distance over a frame is the output of the Color Nuance Descriptor f_{CN}.

b) Multi-Scale Pattern Spectrum [5]: Concepts of granulometry are applied for determining the size distribution of objects in an image without explicitly segmenting each object first. Image objects are treated as particles whose sizes can be established by sifting them through sieves of increasing mesh width and collecting what remains in the sieve after each pass. Due to large patches of uniform color in cartoons, differences between cartoons and natural images in the corresponding pattern spectrums are expected.

Edge-Descriptor: In [3] the Canny and in [5] the Sobel descriptor were implemented for edge detection.

Motion Activity-Descriptor: In [3] and [6] the motion is measured by a pixel-wise color difference on consecutive frames to create maps emphasizing areas of high and low motion. Additionally [6] analyze the derivative of this difference signal. The same features were used for the multi-category case in [28] and [29]. It is more efficient to compute the Motion Activity from the MPEG motion vectors mentioned in [3].

Shot Transition-Descriptor: In [30] the average shot length and the transition type were used as a discriminatory feature to classify cartoons.

Audio-Descriptors: Mel Frequency Cepstrum Coefficients (MFCC) were used in two-category [3] and multi-category experiments [28], [29].

3.2. Our improved Approach

To distinguish between 'cartoon' and 'non-cartoon' videos we analyzed the performance of several combinations of our new developed descriptors and classification methods. The best result was achieved by two HMM's and five visual descriptors: Color Nuance f_{CN} [3], Color Count f_{CC}, Logo Recognition f_{LogR}, Logo Detection f_{LogD} and Dominant Color f_{DC} (Fig. 2).

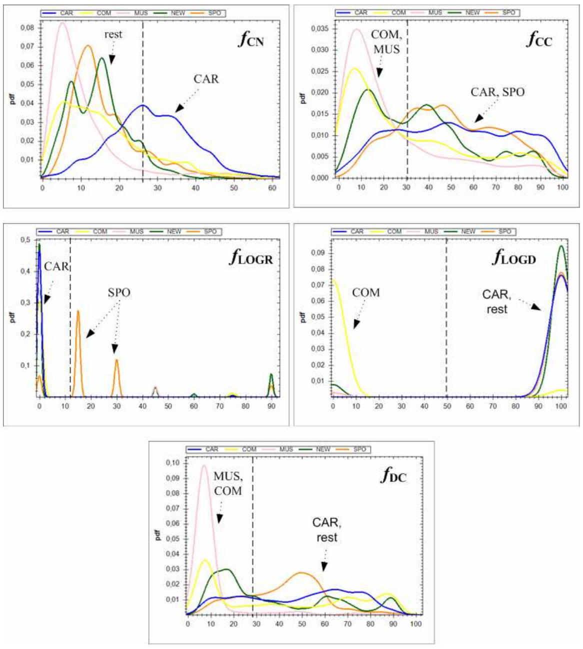

Figure 2. Results (pdf) of selected descriptors

The results (classification accuracy CA) of the tested classification methods within a decision window of $T_{\mathrm{W}}\approx 20$ sec are presented in Table 2.

Table 2. Cartoon detection with different classifiers

Methods	CA
HMM	**95,6%**
Decision Tree	90,3%
Bayes	88,4%
MLP	77,4%

The CA of 95.6% was achieved by a method, using two HMM models, to get a relative probability (Fig.3). One model Θ_{CAR} was designed for the genre cartoon and the other $\Theta_{\underline{\mathrm{CAR}}}$ for non-cartoon.

The models consist of 2 hidden and 6 visible states.

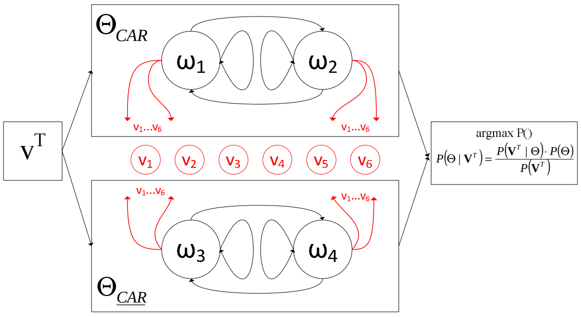

Figure 3. Structure of the cartoon classifier

The 6 visible states $\{v_1, v_2 \ldots v_6\}$ were individually defined by specific combinations and thresholds of the selected descriptors $\{f_{CC}, f_{CN}, f_{DC}, f_{LOGD}$ and $f_{LOGR}\}$.

4. Commercials

Television commercials appear in different forms like spots, exclusive position and sponsoring. Up to present, publications have focused on detecting commercial blocks consisting of several spots.

4.1. Related Work

The content of commercial spots is very diverse. Therefore researchers focus on more general characteristics like separating frames, absence of station-logo and high cut rate according to Table 3:

Table 3: Publications related to commercial detection

Publ.	Genres	Database & Cliplength	Video-features	Audio-features	Classification	Results ($T_{Decision}$)
[7]	COM, COM	5x 20 Clips à 2 min	Separating Block, Static Area (Logo Detection), Hard Cut	/	Decision Tree	CA = 98% (25 s)
[8]	COM, COM	/	Black Frames, Hard Cut Rate	Silence Detection	/	Pure Descriptor Proposals
[9]	COM, COM	6 Clips à ca. 2 h	Logo Detection, Shot Length	/	HMM	CA > 99%
[10]	COM, COM	10 Clips à 10-30 min	Black Frame Detection	Silence Detection	Threshold	Prec = 100% Recall = 87% (90 s)
[11]	COM, COM	10 Clips à 10-30 min	Black Frame Detection	Silence Detection	Threshold	Prec = 76% Recall = 66%
our result	COM, COM	5x 20 Clips à 3 min	Logo Detection, Hard Cut (Freq.)	/	HMM	CA=98% (20 s)

Separating Blocks: [7], [8], [10] and [11] have identified separating blocks, consisting of dark frames between consecutive spots with different methods. [10] and [11] additionally considered silence as an important audio information during separating blocks.

Absent TV-Station Logos: In [7] the absence of TV-station logos during commercial blocks is identified as a characteristic feature. The static area descriptor was developed to detect absent logos in all four corners of a frame.

High Cut Rate: Many commercial spots have a high cut rate with the intention to maximize the visual impact on the consumer. [7] and [11] use this feature and applied corresponding descriptors.

Commercials have also been classified within the multi-category classification [28], [29] and [30] by using the descriptors presented already in section 3.1. Employing more complex features, [32] have used a text and face descriptor to detect combinations prevalent in commercials and three other genres.

4.2. Our improved Approach

To distinguish between 'commercial' and 'non-commercial' we analyzed the performance of several combinations of our new descriptors and classification methods. The best result Table 4 was achieved by two HMM's and two visual descriptors: Logo Detection f_{LOGD}, which extracts the information of logo occurrence and separating blocks simultaneously and Hard Cut Frequency descriptor f_{CFREQ} (Fig. 4).

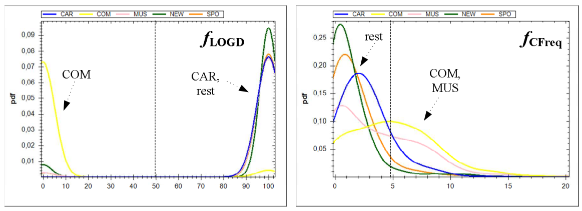

Figure 4. Results (pdf) of selected descriptors

The CA of 98.4% was achieved with two HMM's (Fig.5): one model Θ_{COM} for 'commercial' and the other $\Theta_{\underline{COM}}$ for 'non-commercial'.

Table 4. Commercial detection with different classifiers

Methods	CA
HMM	**98,4%**
Decision tree	98,2%
Bayes	96,1%
MLP	96,1%
SVM	96,1%

The models consist of 2 hidden and 3 visible states.

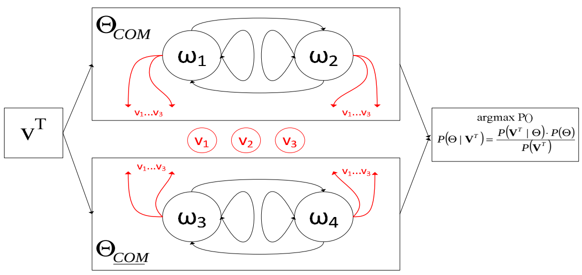

Figure 5. Structure of the commercial classifier

5. Music Videos

Considering the wide spectrum of the music genre, TV and internet are dominated by entertainment music. The most characteristic features of music videos rely on the audio level, as the video content is very diverse.

5.1. Related Work

The main approaches tackle the intra-genre music classification problem using MFCC in [12], [14], [15], [33] and MPEG-7 audio descriptors [2] and [16] according to Table 5.

Table 5. Publications related to intra-music classification

Publ.	Nr. of Genres	Genres	Database & Cliplength	Video-features	Audio-features	Classification	Results ($T_{Decision}$)
[12]	10	Various Music Sub-Gernes	10x 100 Clips à 30 s	/	MFCC, Spec. Centroid, Spec. Rolloff, Spec. Flux, ZCR, Low Energy, DWCH	SVM	CA = 67-83%
[13]	6	Various Music Sub-Gernes	6 Clips à 1 h	/	Zipf Feature, TC, FC, TSC	SVM, MLP	CA = 71%
[14]	4	Classic, Jazz, Pop, Rock	4x 25 Clips à 2.000 frames	/	Beat Spectrum, LCP, ZCR, Spectrum Power, MFCC	SVM	CA = 93%
[15]	4	Classic, Pop, Country, Jazz	4x 50 Clips à 30 s	/	MFCC, LPC, Delta, Acc	HMMs	CA = 89%
[16]	4	Music, Speech, Noise, Others	107 Clips	/	Silence Ratio, Spectral Centroid, Harmonicity, Pitch	Decision Tree	CA = 75-89%
our result	2	MUS, MUS	5x 20 Clips à 3 min	Brightness (Avg.), Scalable Color, Logo Detector	Audio Spectrum Spread	Decision Tree	CA = 96% (20 s)

The inter-genre detection, described in [28], [29] and [30] uses the same descriptors.

5.2. Our improved Approach

In contrast to the intra-genre classification, our interest focuses on 'music video' detection. We use the Audio Spectrum f_{ASP} [2] descriptor on the audio-level as well as the video descriptors Brightness f_{BAVG}, Scalable Color f_{SC} [1] and Logo Detection f_{LOGD} on video-level (Fig. 6).

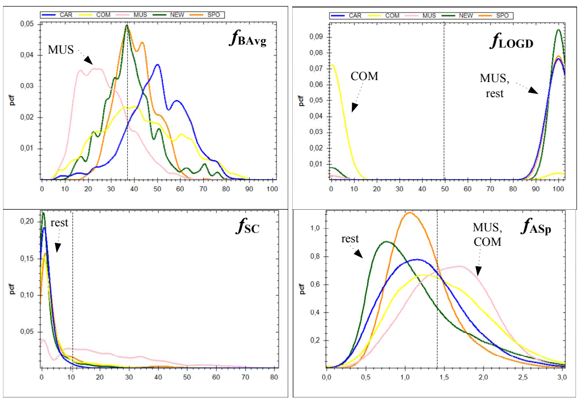

Figure 6. Results (pdf) of selected descriptors

The CAs' of the classifiers are presented in Table 6.

Table 6. Music video detection with different classifiers

Method	CA
Decision Tree	95,5%
HMM	87,3%
SVM	85,7%
Bayes	82,7%
MLP	79,0%

The CA of 95.5% was achieved by a decision tree (ID3) with parameters derived from the pdfs (Fig. 7).

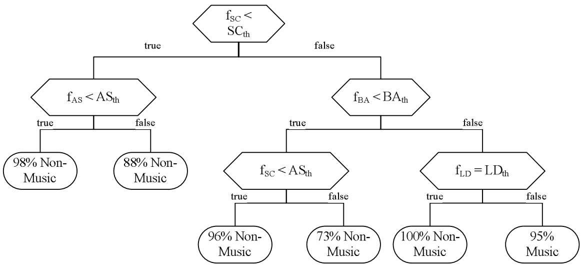

Figure 7. Structure of the music video classifier

6. News

Among news and information broadcasts, publications have focused on detection of newscasts.

6.1. Related Work

Within the scope of inter-genre classification, experiments have been primarily set up for the multi-category detection. The descriptors applied in [28], [29], [30] and [32] have been already explained in sections 3.1 and 4.1.

Anchor shots have a key role in news; therefore their detection is a relevant task for news classification (Table 7). Experimental setups [17]-[21] have focused on anchor shot detection within news shows, but have not been considered in order to separate news from non-news.

Table 7. Publications related to news detection

Publ.	Genres	Database & Cliplength	Video-features	Audio-features	Classification	Results ($T_{Decision}$)
[17]	Various Story Categories	8 Clips à 30 min	Face Detector, Text Detection, Motion, Background Color, Settings Color	Silence, Speech, Music, Noise	HMM, Bayes Net	CA = 98%
[18]	Various Shot Categories	200 Clips à 60-90 s	Caption, Color, Motion	Frame Energy, ZCR, MFCC	SVM	CA = 92%
[19]	Various Shot Categories	218 Clips à 30 min	Anchor Face, Station Logos, Caption Titles	Pitch Jump, Significant Pause, Speech Segments	Maximum Entropy	CA = 91%
[20]	Various Shot Categories	14 Clips à 20 min	Shot Boundary Detection	/	Fuzzy	Prec = 98% Recall = 96%
[21]	5 Shot-Categories	75 min of 468 Shots	Cut Detection, Facial Region, Skin-colored Region, Template Matching, Lip Movement, OCR	Speech Recognition	Rule	Success Rate > 75%
our result	NEWS, NEWS	5x 20 Clips à 3 min	Logo Detector, Marquee	LEF, HiZCR, Spectrum Spread Modulation	Decision Tree	CA = 92% (20 s)

6.2. Our improved Approach

To distinguish between the 'news' and 'non-news' videos we analyzed again the performance of several combinations of our new developed descriptors and classification methods. The preferred descriptors are shown in Fig. 8: Logo Detection f_{LOGD}, Marquee f_{MARQ}, Audio Low Energy f_{LEF}, Audio High Zero Crossing Rate f_{HZCR} and Audio Spectrum Modulation f_{SMOD}.

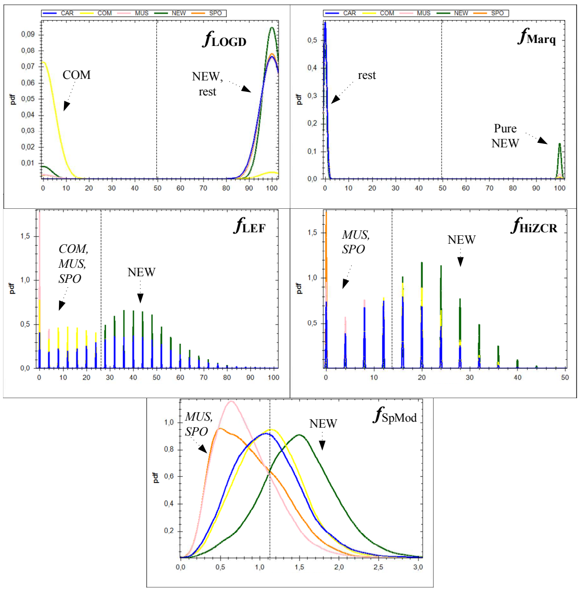

Figure 8. Results (pdf) of selected descriptors

Whereas the main research [17]-[21] focus on intra-genre news detection using shot segmentation; our approach tackles the inter-genre classification problem. The classification accuracies are presented in Table 8.

Table 8. News detection with different classifiers

Method	CA
Decision Tree	**91,9%**
Bayes	86,7%
MLP	86,0%
HMM	70,0%

The best CA of 91.9% was achieved by a decision tree (ID3) with parameters derived from the pdfs.

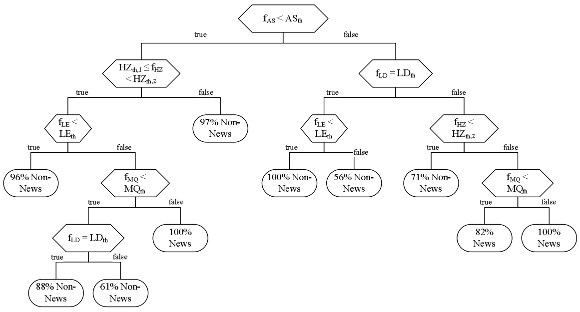

Figure 9. Structure of the news classifier

7. Sports

Sport is a very diverse domain, containing sub-genres like basketball, football, motor sports, soccer and tennis e.g. For this reason one approach tries to differentiate between different sports types within the sport domain, whereas a second approach classifies sport as part of the multi-category-problem. The multi-category problem can be reduced to sport/non-sport detection [22]-[26] by using the features: motion, dominant color, field-lines and shot length according to Table 9.

Table 9. Publications related to sport detection

Publ.	Genres	Database & Cliplength	Video-features	Audio-features	Classification	Results ($T_{Decision}$)
[22]	Fieldsport-Events	100 h	Hard Cut, Close Up, Crowd Image, Scoreboard, Motion Activity, Field Lines	Speech Band Audio Activity	SVM	Max. ERR = 97% (25 s)
[23]	SOC: Play-, Break-Detektion	20 Clips à 2-10 min	Color (Field Area, Horizontal Variance, Vertical Variance), Motion Intensity	/	HMMs Hierarchy	CA = 83%
[24]	SOC: COM, Play-, Break-Detection	7-46 min	Dominant Color-Ratio (Green), Camera Views	Speech, Music	Rules	CA = 94%
[25]	TEN-Syntax	8 Clips, 5 h	Shot length, Dominant Color, Relative Player Position	Speech, Applause, Ball Hits, Noise, Music	HMMs Hierarchy	CA = 86%
[26]	HOC, SWIM, TRA, YAC	61 Shots	37 Cues (e.g. Players, Ball, Face, Camera Motion, etc.)	/	Tree (C4.5), 5 HMMs	CA = 83%
our result	SPO, <u>SPO</u>	5x 20 Clips à 3 min	Motion Activity (Intensity, Acceleration), StaticArea, Color (Sportfield), Spectator, Logo-Detector	/	Bayes	CA = 95% (20 s)

BASE = Baseball, BASK = Basketball, CYCL = Cycling, FOOT = American Football, ICEH = Icehockey, SOCC = Soccer , SWIM = Swimming TENN = Tennis, VOLL = Volleyball, YACH = Sailing, EVEN = Events, SPO = Sport, NEWS = News, MOV = Film, COM = Commercial, MUS = Music video

7.1. Our improved Approach

To distinguish between 'sport' and 'non-sport' videos we used the following descriptors: Spectator f_{SPEC}, Sportfield f_{SF}, Static Area f_{STATA}, Logo Recognition f_{LOGR}, Motion Activity f_{MA} and Acceleration f_{MOTACC}.

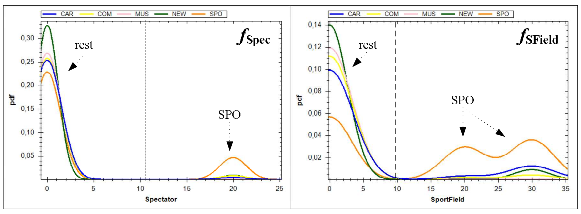

Figure 10. Results (pdf) of selected descriptors

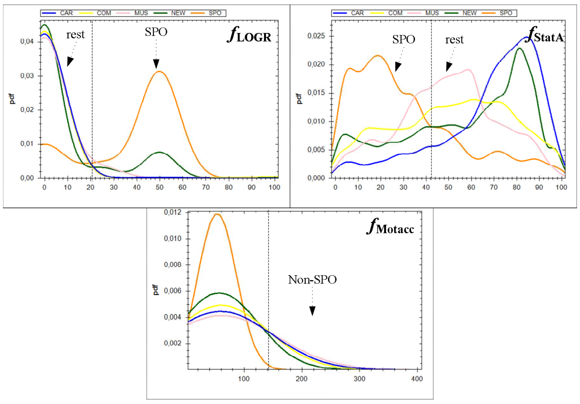

Figure 10. Results of selected descriptors (cont.)

The classification accuracies of the analyzed methods are presented in Table 10.

Table 10. Sport detection with different classifiers

Method	CA
Bayes	**95,2%**
Decision Tree	93,5%
MLP	72,9%
SVM	79,8%

We achieved the best results using the Bayes' Theorem according to Fig. 10. The a priori probabilities were assumed as 1/5 for 'sport' and 4/5 for 'non-sport'.

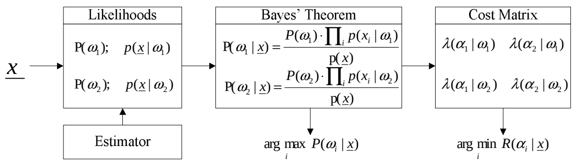

Figure 11. Structure of the sport classifier

8. Multi-Genre

Researchers have applied various classification methods in their multi-genre approaches. A review of some publications is shown in Table 11.

Table 11. Publications related to multi-genre detection

Publ.	Nr. of Genres	Genres	Database & Cliplength	Video-features	Audio-features	Classification	Results ($T_{Decision}$)
[27]	5	CAR, COM, MUS, NEW, SPO	5x 1 h à 5 min	Scalable Color, Color Layout, Homogeneous Texture	14 MFCC	GMM	CA = 87%
[28]	5	CAR, COM, MUS, NEW, SPO	5x 1 h à 5 min	Motion Activity	MFCC	GMM	ERR = 10% (20 s)
[29]	3	CAR, NEW, SPO	18 Clips à 50 s	Motion (Camera, Object)	/	GMM	EER = 6% (30 s)
[30]	5	CAR, COM, MUS, NEW, SPO	720 à (40 s) 480 à (60 s) 360 à (80 s)	Average Shot Length, Transition Type, Camera Motion, Luminance Variance, Color Histograms, Saturation	/	Tree (C4.5)	CA = 82% (80 s), 80% (40 s)
[31]	4	CAR, COM, NEW, SPO1, SPO2	5x2 Clips à 4 min	Cut Detection, Motion (Camera, Object, Energy), Logo Recognition	Audio FFT (speech, music, silence, noise)	/	Pure Descriptor Performance
our result	5	CAR, COM, MUS, NEW, SPO	5x 20 Clips à 3 min	Logo (Detection, Recognition), DCD Color Nuance, Color Count	/	Bayes, HMMs, Decision Trees	CA = 95% (Clip), 88% (20 s)

The here applied descriptors have been described in the respective five two-category Ω_2={c, $\underline{c}$} cases.

9.1. Our improved Approach

The decisions of our five two-category detectors are combined by different levels [34], diverse structures and several rules (Table 12).

Table 12. Multi-Genre detection in different combinations

Structure	Level	Rule / Method	CA
parallel	**opinion**	**bayes**	**88.4%**
parallel	opinion	decision tree	82%
parallel	abstract	weight. voting	82.3%
parallel	opinion	product	88.3%
parallel	opinion	sum	74.8%
serial	abs. + opinion	sum	80.8%
serial	abstract	rejection class	ERR=4.8%

We achieved the best result for multi-genre detection with a parallel structure, in opinion level using a Bayesian post-classifier.

9. Experiments

The experiments were carried out on a representative database of 100 mpeg-2 video sequences in total of 300 min of recordings; 20 sequences per genre gathered from popular TV broadcasting in Europe and Internet. The different test-sets (50 video sequences) for the following experiments has been chosen randomly. Our goal was to get a result within a decision window of T_W≈20 sec.

9.1. Experimental results

The results demonstrate high identification rates of more than 90% for each genre (Table 13 – Table 18).

Table 13. Confusion matrix for cartoon classifier

true / pred	***Non-Cartoon***	***Cartoon***	***Recall***
Cartoon	17	46	73%
Commercial	84	0	100%
Music	71	0	100%
News	107	1	99,1%
Sport	84	0	100%
Precision	95,3%	97,9%	**CA = 95,6%**

Table 14. Confusion matrix of the commercial classifier

true / pred	***Non-Commercial***	***Commercial***	***Recall***
Cartoon	90	0	100%
Commercial	5	79	94%
Music	70	1	98,6%
News	115	1	99,1%
Sport	84	0	100%
Precision	98,6%	97,5%	**CA = 98,4%**

Table 15. Confusion matrix of the music video classifier

true / pred	***Non-Music***	***Music***	***Recall***
Cartoon	65	0	100%
Commercial	84	0	100%
Music	15	81	84,4%
News	111	5	95,7%
Sport	84	0	100%
Precision	95,8%	94,2%	**CA = 95,5%**

Table 16. Confusion matrix of the news classifier

true / pred	***Non-News***	***News***	***Recall***
Cartoon	86	8	91,5%
Commercial	83	1	98,8%
Music	90	6	93,8%
News	23	93	80,2%
Sport	81	0	100%
Precision	93,7%	86,1%	**CA = 91,9%**

Table 17. Confusion matrix of the sport classifier

true / pred	***Non-Sport***	***Sport***	***Recall***
Cartoon	69	1	98,6%
Commercial	41	0	100%
Music	79	0	100%
News	111	3	97,4%
Sport	15	76	83,5%
Precision	95,2%	95%	**CA = 95,2%**

Table 18. Confusion matrix of the multi genre classifier

true/pred	***CAR***	***COM***	***MUS***	***NEW***	***SPO***	***Recall***
CAR	136	0	1	14	3	88,3%
COM	0	158	0	2	3	96,9%
MUS	13	2	136	0	16	81,4%
NEW	20	5	5	234	11	85,1%
SPO	8	1	1	4	164	92,1%
Precis.	76,8%	95,2%	95,1%	92,1%	83,3%	**CA 88,4%**

With our current non-optimized 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.

10. Summary & Conclusion

The intention of this paper was to give an overview of approaches to the video-genre classification problem, particularly inter-genre classification which means analyzing a video and assigning automatically different genre labels such as cartoon, commercial, music, news and sport. This problem has been solved by combining various new audio-visual features and classification methods. The different experimental set-ups and corresponding results are summarized for each genre. This technology is interesting in connection with personal video recorder and for the automatic classification of audio-visual data for Internet and search engines.

10. References

[1] T. Sikora, P. Salembier and B.S. Manjunath, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley LTD, ISBN 0471486787, 2002.

[2] H.G. Kim, N. Moreau and Thomas Sikora, *Introduction to MPEG-7 Audio: Content Indexing and Retrieval*, Wiley & Sons, ISBN-10: 047009334.

[3] R. Glasberg, A. Samour, K. Elazouzi, and T. Sikora, “Cartoon recognition using video & audio descriptors,” in Proceedings of the European Signal Processing Conference (EUSIPCO), 2005.

[4] B. Ionescu, P. Lambert, D. Coquin, L. Darlea, “Color-based semantic characterization of cartoons,” Signals, Circuits and Systems, 2005. ISSCS 2005. International Symposium on Volume 1, 14-15 July 2005 Page(s):223 - 226 Vol. 1 .

[5] T. I. Ianeva, A. P. de Vries, and H. Rohrig, “Detecting cartoons: A case study in automatic video-genre classification,” in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), vol. 1, Jul. 2003, pp. 449–452.

[6] M. Roach, J. S. Mason, and M. Pawlewski, “Motion-based classification of cartoons,” in Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing, Mai 2001, pp. 146–149.

[7] R. Glasberg, C. Tas, and T. Sikora, “Recognizing commercials in realtime using three visual descriptors and a decision-tree,” in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), July 2006.

[8] B. Satterwhite and O. Marques, “Automatic detection of tv commercials,” in IEEE Potentials Magazine, May 2004, pp. 9–12.

[9] A. Albiol, Fullà, A. Albiol and L. Torres, “ Detection of TV commercials,” Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on Volume 3, 17-21 May 2004 Page(s):iii - 541-4 vol.3.

[10] D. A. Sadlier, S. Marlow, D. N. O’Connor, and N. Murphy, “Automatic TV advertisement detection from MPEG bitstream,” Journal of the Pattern Recognition Society, vol. 35, no. 12, pp. 2–15, December 2002.

[11] S. Marlow, D. A. Sadlier, K. McGeough, N. O’Connor, and N. Murphy, “Audio and video processing for automatic TV advertisement detection,” in Proceedings of the IEEE Symposium on Computers and Communications (ISSC), 2001.

[12] T. Li and M. Ogihara, “Music genre classification with taxonomy,” Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on Volume 5, 18-23 March 2005 Page(s): v/197 - v/200.

[13] E. Dellandrea, H. Harb and L. Chen, “Zipf, Neural Networks and SVM for Musical Genre Classification,” Signal Processing and Information Technology, 2005. Proceedings of the Fifth IEEE International Symposium on 18-21 Dec. 2005 Page(s):57 – 62.

[14] C. Xu, N. C. Maddage, X. Shao, F. Cao and Q. Tian, “Musical Genre Classification Using Support Vector Machines”, Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on Volume 5, 2003 Page(s):V - 429-32 vol.5

[15] X. Shao, C. Xu and M.S. Kankanhalli, „Unsupervised classification of music genre using hidden Markov model,” Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on Volume 3, 27-30 June 2004 Page(s):2023 - 2026 Vol.3.

[16] G. Lu and T. Hankinson, “A technique towards automatic audio classification and retrieval,” Signal Processing Proceedings, 1998. ICSP '98. 1998 Fourth International Conference on Volume 2, 12-16 Oct. 1998 Page(s):1142 - 1145 vol.2.

[17] F. Colace, P. Foggia and G. Percannella, “A Probabilistic Framework for TV-News Stories Detection and Classification,” Proceedings of the IEEE International Conference on Multimedia and Expo (ICME).

[18] W. Lie and C. Su, „News Video Classification Based on Multi-modal Information Fusion,” Image Processing, 2005. ICIP 2005. IEEE International Conference on Volume 1, 11-14 Sept. 2005 Page(s): I - 1213-16.

[19] W. Hsu, L. Kennedy, C.W. Huang, S.F. Chang, C.Y. Lin and G. Iyengar, “News Video Story Segmentation Using Fusion of Multi-Level Multi-Modal Features in TRECVID 2003,” Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[20] X. Gao and X. Tang, “Unsupervised Video-Shot Segmentation and Model-Free Anchorperson Detection for News Video Story Parsing,” Circuits and Systems for Video Technology, IEEE Transactions on Volume 12, Issue 9, Sep 2002 Page(s):765 – 776.

[21] I. Ide, K. Yamamoto, R. Hamada and H. Tanaka, “An Automatic Video Indexing Method Based on Shot Classification,” Systems and Computers in Japan, Vol. 32, No. 9, 2001.

[22] D.A. Sadlier and N. O'Connor, “Event Detection in Field Sports Video Using Audio-Visual Features and a Support Vector Machine,” Transactions on Circuits and Systems for Video Technology, Vol.15, 2005.

[23] F. Wang, Y.F. Ma, H.J. Zhang and J.T. Li, „A Generic Framework for Semantic Sports Video Analysis Using Dynamic Bayesian Networks,” Proceedings of the 11th International Multimedia Modelling Conference (MMM), 2005.

[24] J. Chen, Y. Li, S. Lao and L. Wu, “A Unified Framework for Semantic Content Analysis in Sports Video,“ Proceedings of the 2nd International Conference on Information Technology for Application (ICITA), 2004.

[25] P. Gros, E. Kijak and G. Gravier, “Automatic Video Structuring Based on HMMs and Audio Visual Integration,” Proceedings of the Second International Symposium on Image/Video Communications over Fixed and Mobile Networks, 2004.

[26] E. Jaser, J. Kittler and W. Christmas, “Hierarchical Decision Making Scheme for Sports Video Categorisation with Temporal Post-Processing,“ Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2004 .

[27] L.Q. Xu and Y. Li, “Video Classification Using Spatial-Temporal Features And PCA,” Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 2003.

[28] M. Roach, J.S. Mason and L.Q. Xu, “Video Genre Verification using both Acoustic and Visual Modes,” Proceedings of the IEEE Workshop on Multimedia Signal Processing, 2002.

[29] M. Roach, J.S. Mason and M. Pawlewski, “Video Genre Classification Using Dynamics,” Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2001.

[30] B.T. Truong, S. Venkatesh and C. Dorai, “Automatic Genre Identification for Content-Based Video Categorization,” Proceedings of the 15th International Conference on Pattern Recognition, 2000.

[31] S. Fischer, R. Lienhart and W. Effelsberg, „Automatic Recognition of Film Genres“, Reihe Informatik 6/95, Universität Mannheim.

[32] N. Dimitrova, L. Agnihotri and G. Wei, “Video Classification Based On HMM Using Text And Faces,” Proceedings of the European Conference on Signal Processing, 2000.

[33] H.-G. Kim, N. Moreau, and T. Sikora, “Audio classification based on MPEG-7 spectral basis representations,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 5, pp. 716–725, 2004.

[34] A. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: a review,” in IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 22, No.1, January 2000.