CARTOON-RECOGNITION USING VISUAL-DESCRIPTORS AND A MULTILAYER-PERCEPTRON

Ronald Glasberg, Khalid Elazouzi and Thomas Sikora
glasberg@nue.tu-berlin.de, elazouzi@cs.tu-berlin.de, sikora@nue.tu-berlin.de

Communication Systems Group
Technical University Berlin
10587 Berlin, Germany

ABSTRACT

We present a new approach for classifying mpeg-2 video sequences as 'cartoon' or 'non-cartoon' by analyzing specific color, texture and motion features of consecutive frames in real-time. This is part of the well-known video-genre-classification problem, where popular TV-broadcast genres like cartoon, commercial, music, news and sports are studied. Such applications have also been discussed in the context of MPEG-7 [12]. In our method the extracted features from the visual descriptors are non-linear weighted with a sigmoid-function and afterwards combined using a multilayered perceptron to produce a reliable recognition. The results demonstrate a high identification rate based on a large collection of 200 representative video sequences (40 cartoons and 4*40 non-cartoons) gathered from free digital TV-broadcasting in Germany.

1. INTRODUCTION

With the advent of digital TV-broadcasts presenting more than hundred of channels at a time, the need for a user-friendly TV-program selection is growing. Unlike the present TV, a new system should enable users to access programs clustered by genres. The main goal of our research is therefore the classification of an mpeg-2 video-stream in real-time into genres at the highest level. Our interest is in identifying cartoons in broadcasting TV material.

2. RELATED WORK

The recent approaches addressing cartoon-detection and video-classification are listed in [1-3] and [4-9] respectively. These methods extract a number of different low-level features, from which the analysis is made to build so-called signatures to describe a certain video class. Roach et al. published an approach for the classification of sequences as cartoons using one descriptor [2] extracting a motion-feature on a database of 8 cartoons and 20 non-cartoon sequences all together of 20 minutes. Athitsos et al. [3] emphasized the basic characteristics of cartoons and implemented nine color descriptors in order to distinguish photographs and graphics on the www on a database of 1200 samples. In the same fashion, but with more insight into the basic characteristics of cartoons, Ianeva et al. [1] implemented six descriptors. It is difficult to predict how the described systems would perform on a test set of mpeg-2 streams.
We use a database of 200 representative video sequences (100 sequences for training and 100 as testing-data, e.g. cartoons with dark frames, commercials with animated cartoon sequences) and five new/ modified visual descriptors and a MLP.

3. VIDEO-GENRE-CLASSIFICATION PROCESS

In a first step each mpeg-2 stream is divided into the visual and acoustical data.

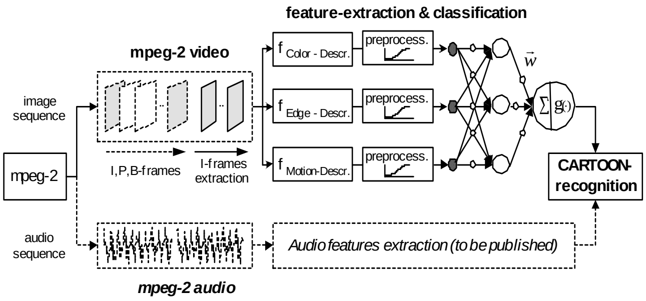

Figure 1: Process involved computing the feature vectors

We consider in this paper only consecutive I-frames of the image sequence and save them temporarily for processing.

3.1. Visual-Descriptors

By looking at frames from cartoon and non-cartoon sequences we observed basic differences between them, like the appearance of bright, highly saturated colors, areas of uniform color, few sharp edges and a low motion activity. Based on these observations, we designed five descriptors to transform our data into feature vectors.

3.1.1. Brightness-Descriptor family

We take each frame and determine the average brightness as well as the amount of pixels with a level higher than a threshold Th_L. These values are averaged over a window N resulting in f_1 and f_2 respectively.

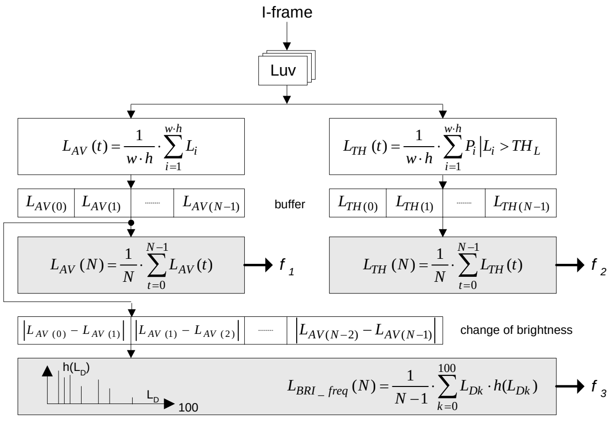

Figure 2: Block diagram of the brightness-descriptor

Our interest is also focused on the change of brightness for consecutive frames; therefore we developed $L_{Bri\text{-}freq}$ with the output f_3.

3.1.2. Saturation-Descriptor family

The saturation-descriptor uses the HSV-Space. Similar to the brightness-descriptor we determine the change of saturation f_5 for consecutive frames.

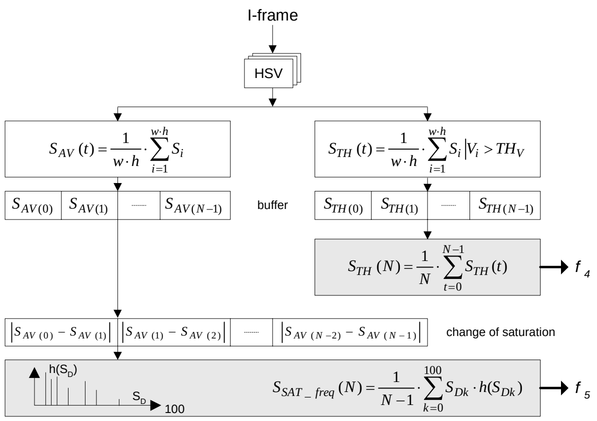

Figure 3: Block diagram of the saturation-descriptor

For the threshold f_4 we consider only the saturation with a brightness-value higher than a threshold Th_V.

3.1.3. Color Nuance Descriptor

We determine the mean color distance of each pixel with his adjacent eight neighbors (except border); calculate the mean value for the frame, over a window N and obtain f_6.

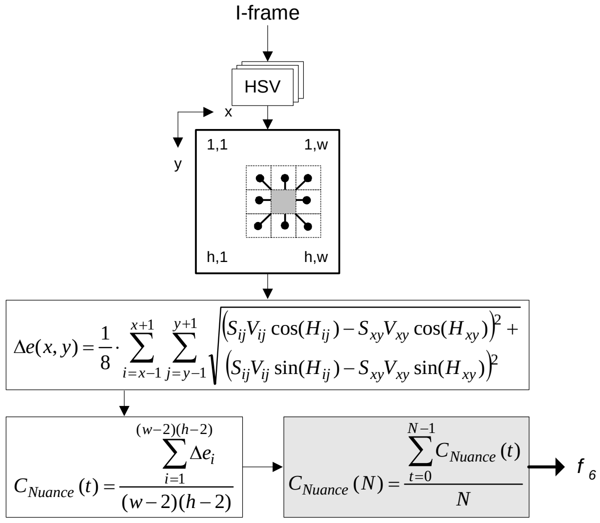

Figure 4: Block diagram of the color-nuance-descriptor

3.1.4. Edge-Descriptor

We detected, that Cartoons have in comparison to the other mentioned genres in general less sharp edges. Therefore we implemented the Canny detector [10]. In our experimental tool the result is labeled with f_7.

3.1.5. Motion-Descriptor family

We implemented one descriptor [11] which extracts the motion-activity information included in the mpeg-2 stream as well as one descriptor extracting the motion information f_8 of a video-stream similar to [2].

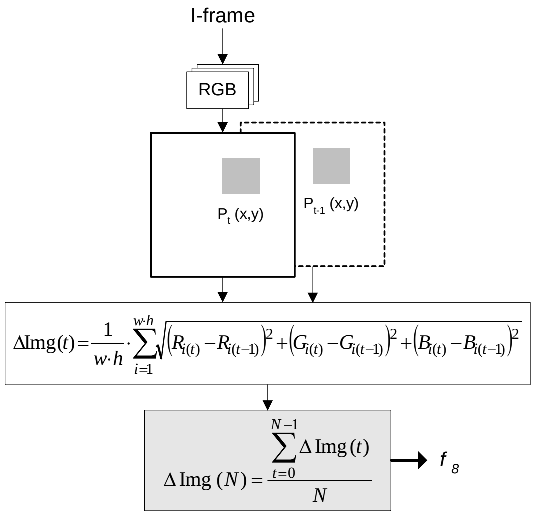

Figure 5: Block diagram of the motion-descriptor

4. EXPERIMENTS

The experiments were carried out on a representative large collection in total of 400 min of recordings; 40 cartoons and 4*40 non-cartoons (commercial, music, news and sport) of 2 minutes' each gathered from popular TV in Germany (ARD, ZDF, BBC, RTL, SAT1, VIVA..). 50% of the data was for training and the rest for testing. We extracted consecutive I-frames and scaled them down to a resolution of 90*72 pixels. The number of considered frames in the preprocessing window is N=50.

4.1. Experimental Results

Figure 6 shows the 8 pdf's probability density functions for descriptor $f_1 \dots f_8$.

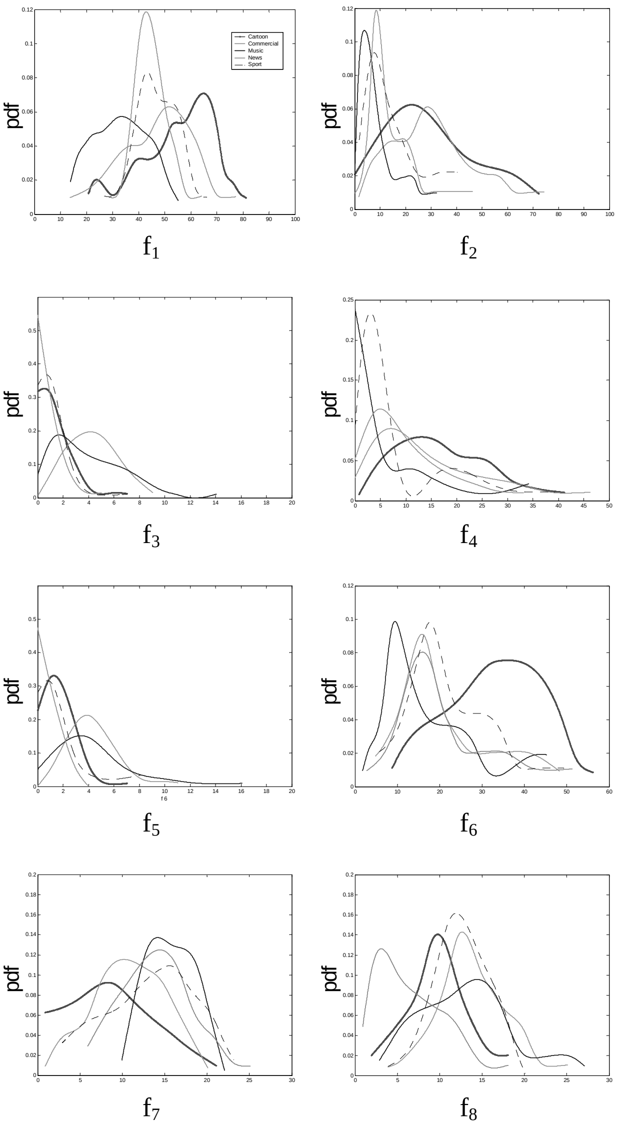

Figure 6: Results of the descriptors f_1-f_8 for each genre

The recognition of cartoons with the extracted feature of a single descriptor is obviously not sufficient. In order to achieve high identification rates, we used the output of each descriptor as input for the multilayered perceptron. The adjustments of the net-parameters were accordingly to the performance of descriptors f_1-f_8 from Figure 6.

The results of windows N=50 were averaged over the whole length of each video sequence. Fig. 7 depicts the detection rates for the 20 video sequences of each genre used in our experiments from the test-database.

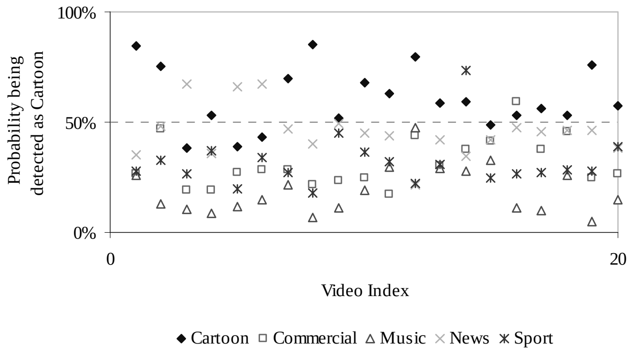

Figure 7: Probability for videos being detected as Cartoon

In our experiment 16 videos from the cartoon-genre were classified as cartoons. The remaining 4 videos were very close to the cartoon-detection threshold of 50%. Those sequences included long dark shots and many non-uniform color areas. It is obvious, that detection of any genre from dark image sequences is highly unreliable. An increase of detection rate will be possible by taking the unreliability into account in the averaging process of the windows for each sequence.

It is interesting to note that 20 videos from the music genre, 19 videos from the commercial and sports genre and 17 videos from the news genre were detected as 'non-cartoon'. From Fig. 7 it can be depicted, that it is possible to detect music videos using our approach as well, as they show in general a probability rate of less then 20%.

The table below shows a summary for the experiment results.

Genre	Cartoon
Cartoon	80%
Commercial	5%
Music	0
News	15%
Sports	5%

Table 1: Classification accuracy on the test-database of 100 mpeg-2 streams

5. SUMMARY & CONCLUSION

We have presented in this paper a nonlinear approach for the detection of cartoons. Three key contributions have been made. We started with the development of new or modified visual descriptors. Second, we considered the temporal features and used a nonlinear sigmoid function. Third, we developed and applied a MLP to combine the results of the visual descriptors, deriving a probability rate for a video being a ‘cartoon’ or ‘non-cartoon’.

A video database containing five popular genres namely cartoon, commercial, music, news and sports has been used. An average correct classification rate of 80% for cartoon-videos detected as a ‘cartoon’ and more than 85% for the other genres detected as a ‘non-cartoon’ has been achieved, using a window of N=50 frames and averaging the results over the sequence.

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.
We took great care to perform experiments on a large and balanced database of video from real, compressed broadcast material. From our results and database it is not obvious, that the descriptor used in [2], same as our descriptor f_8, by itself will be sufficient for classification.

6. REFERENCES

[1] T.I. Ianeva, A.P. de Vries and H. Röhrig, “Detecting cartoons: A case study in video-genre classification”, Proceedings ICME Multimedia and Expo, volume: 1, pp. 449-452, 2003.

[2] M. Roach, J.S. Mason and M. Pawlewski, “Motion-based classification of cartoons”, International Symposium on Intelligent Multimedia, pp. 146 – 149, 2001.

[3] V. Athitsos, M.J. Swain and C. Frankel, “Distinguishing photographs and graphics on the world wide web”, Proceedings IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 10 – 17, 1997.

[4] L.Q. Xu and Y. Li, “Video classification using spatial-temporal features and pca”, Proceedings ICME Multimedia and Expo, volume: 3, pp. 485-8, 2003.

[5] 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.

[6] M. Roach, J. Mason, N. Evans and L.-Q. Xu, “Recent trends in video analysis: a taxonomy of video classification problems”, Proceedings Internet and Multimedia Systems and Applications, pp. 348 – 354, 2002.

[7] M. Roach, J. Mason and M. Pawlewski, “Video genre classification using dynamics”, Proceedings Acoustics, Speech, and Signal Processing, volume: 3, pp.1557 – 1560, 2001.

[8] 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.

[9] S. Fischer, R. Lienhart and W. Effelsberg, “Automatic recognition of film genres”, third ACM International Multimedia Conference and Exhibition, pp. 295-304, 1995.

[10] J. Canny, “A computational Approach for Edge Detection”, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6 pp. 679-698, 1986.

[11] H. Krambeck, R. Glasberg and T. Sikora, “Development of an Analyzer for Video-Genre-Classification of mpeg-2 streams in digital video broadcasting”, TU-Berlin, *student research project*, 2004.

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