CLASSIFICATION OF ANIMATED VIDEO GENRE USING COLOR AND TEMPORAL INFORMATION

Bogdan Ionescu[1] and Patrick Lambert[2]

We address a particular case of video genre classification, namely the classification of animated movies. We propose two categories of content descriptors: temporal and color based, which are adapted to this particular task. Temporal descriptors, like rhythm or action, are quantifying the perception of the action content at different levels. Color descriptors are determined using color perception which is quantified in terms of statistics of color distribution, elementary hues, color properties and color relationship. Experimental tests conducted on more than 159 hours of video footage and various classification schemes show the efficiency of this approach. Despite the high diversity of the video material, the proposed descriptors are able to provide an average global correct classification up to 92.7%.

Keywords: animated genre classification, action content, color properties.

1. Introductions

Accessing multimedia information or ”content” is now part of our daily routine. The actual challenge is how to make useful this information and retrieve relevant content. We reached the point where a device should find multimedia content for us just as another person would do. To this end, significant efforts are currently made to develop innovative automatic content-based indexing techniques. Of particular interest is the automatic cataloging of video footage into some predefined semantic categories. This can be performed *globally*, by classifying videos into one of several main genres, e.g. cartoons, music, news, sports. Also, sub-genres can be involved, e.g. identifying specific types of sports (football, hockey, etc.), movies (drama, thriller, etc.), and so on. Another solution aims to classify movie content *locally*, thus considering video segments and specific concepts, e.g. outdoor, action, violence, etc.

[1]Lecturer, LAPI - University ”Politehnica” of Bucharest, Romania, e-mail: `bionescu@alpha.imag.pub.ro`

[2]Professor, LISTIC - Polytech Annecy-Chambery, France, e-mail: `patrick.lambert@univ-savoie.fr`

In this paper we address the global classification of a particular genre, namely the animated movies. The animated movie industry witnessed nowadays a spectacular development and gain in popularity: abundance of entertainment cartoon movies, festivals and expo, e.g. France - Annecy International Animated Film Festival, Canada - Ottawa International Animation Festival, Portugal - CINANIMA International Animation Film Festival, etc. Animated movies now target equally children and adults, becoming a distinctive industry similar to the artistic movies.

In the context of the automatic content-based retrieval, a common task related to this field is the automatic *selection of the "animated" content from other genres*. Regardless the approach, the main challenge is to derive attributes which are discriminant enough to distinguish between genres while maintaining a reduced dimensionality of the feature space. Several approaches have been investigated in the literature.

One approach is to address the classification at *image level*. For instance, [2] emphasizes the basic characteristics of cartoons and uses nine color descriptors to distinguish between photographs and graphics over the World Wide Web. Another example is the approach in [3]. It uses Support Vector Machines (SVM) with several image descriptors, i.e. saturation and brightness information, color histograms, edge information, compression ratio and pattern spectrum to label individual video frames as "cartoon" or "photographic". Authors announce correct classification ratios around 94% when tested on more than 24,000 static images. However, the main limitation of this approach is in its static nature, video specific dynamic information being disregarded.

Another category of approaches (that constitute the subject of this work) perform the classification at *sequence level*, e.g. [4] discusses an uni-modal approach and testes the prospective potential of motion information to cartoon classification. However, experimental validation was performed on a very limited data set, only 8 cartoon and 20 non cartoon sequences, making difficult to predict how the method will perform on a wider database. A two-modal approach is proposed in [5] and cartoon classification is performed using a multilayered perceptron with both visual (brightness, saturation, color hue, edge information, motion) and audio descriptors (MFCC descriptors). Tests were performed on a bit larger database containing 100 sequences (20 sequences of each genre: cartoons, commercials, music, news and sports) and classification accuracy is around 90%. Another example is the approach in [6] which uses eight human inspired MPEG-7 visual descriptors and a SVM scheme with active relevance feedback.

Other methods are addressing the *video genre classification*, which includes the case of cartoon movies. A state-of-the art is available in [8]. For instance, [9] proposes a truly multi-modal approach which combines several types of content descriptors. Features are extracted from four informative sources, which include visual-perceptual information (color, texture and motion), structural information (shot length, shot distribution, shot rhythm, shot clusters duration and saturation), cognitive information (face properties, such as number, positions and dimensions) and aural information (transcribed text, sound characteristics). These pieces of information are used to train a parallel neural network system and provide a maximum accuracy rate up to 95% in distinguish between seven video genres (including cartoons): football, cartoons, music, weather forecast, newscast, talk shows and commercials. However, these techniques, in general, are not focusing on the retrieval of animated content. They are limited to use "all purpose" content descriptors which work well with all genres, but not specifically with animated content.

This short overview of the literature shows in general a lack of dedicated approaches, the few existing ones being limited to address the particular case of classic cartoons (paper drawings) and to use more or less general purpose descriptors. The remainder of this paper is organized as follows: Section **2** highlights the contribution of our work, Section **3** and Section **4** deal with feature extraction: temporal information and color properties. Experimental results are presented in Section **5** while Section **6** presents the conclusions and discusses future work.

2. The proposed approach

The first limitation of the existing approaches is in the targeted genre which is exclusively the cartoon genre. In this paper we extend the classification by addressing, generically, the animated movies and thus including equally cartoons and artistic animated movies. Artistic animated movies, less common than cartoons, but with an increasing popularity, are usually short animated clips having artistic connotations. Contrary to cartoons, artistic animated movies are produced using a high variety of techniques and use artistic concepts (see CITIA [10] and [7]). Basically, any "un-natural" (e.g. using artificially created content) video production can be assigned to the animated category. Some examples are depicted in Fig. 1.

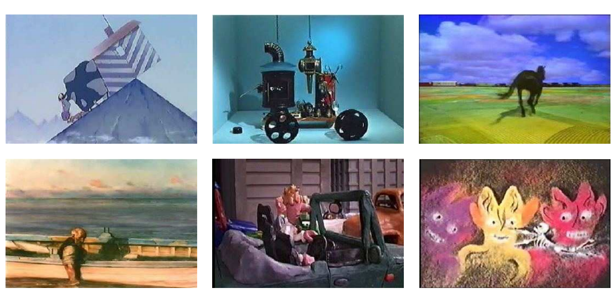

Fig. 1. Various animation techniques (source [10]).

Due to their distinctive creation process and particular contents, animated movies often require a different processing approach than natural movies.

Most of the existing approaches are proposing generic content descriptors, like image-based descriptors: e.g. saturation, brightness, contours or temporal-based: motion, sound, etc., which are not particularly addressing the properties of this genre. For instance, motion is usually discontinuous with animated movies and sometimes impossible to estimate, many artistic movies are without sound or commentaries, etc.

To address these issues, we propose two categories of content descriptors, namely: *temporal-based* (animated movies usually have a different visual rhythm or action content) and *color-based* (color distribution is always specific), which are adapted to the animation content. For the *temporal descriptors*, e.g. rhythm, action, user experiments have been conducted on animated movies to quantify the perception of the action content at different levels. Temporal information is quantified in terms of visual rhythm, action content and amount of gradual transitions. On the other hand, the *color descriptors* have been validated on the semantic analysis of artistic animated movies [7] and have the advantage of capturing also the temporal information (global descriptors). Using a color naming system, color perception is quantified in terms of statistics of color distribution, elementary hues, color properties (e.g. amount of light colors, cold colors, etc.) and relationship of adjacency and complementarity. This work is an extension of the preliminary study proposed in [1].

3. Action description

The first feature set aims to capture the movie's temporal structure in terms of *visual rhythm*, *action content* and *amount of gradual video transitions*, as these parameters are strongly related to movie contents. To do so, first we perform a temporal segmentation, which roughly means parsing the movie intro shots by means of detecting the video transitions. We detect cuts and two of the most frequent gradual transitions, i.e. fades and dissolves. To favor the animated movies, we use specially adapted algorithms: cut detection is performed using the histogram-based approach proposed in [11], while dissolve and fade detection are carried out using the analysis of fading-in and fading-out pixels proposed in [14] and an adaptation of the pixel-level statistical approach proposed in [13], respectively. Further, we determine the following parameters:

Rhythm. To capture the movie's tempo of visual change, we compute the relative number of shot changes occurring within a time interval of $T = 5s$, denoted ζ_T. Then, the rhythm is defined as the movie average shot change ratio, $\bar{v}_T = E\{\zeta_T\}$. Defined in this way, $\bar{v}_T$ represents the average number of shot changes over the time interval T for the entire movie, being a measure of the movie global tempo. High values of $\bar{v}_T$ indicate a movie with a general high change ratio, while small values correspond typically to movies with predominant long and static shots (a reduced number of scenes).

Action. To determine the following parameters, we use a relatively confirmed assumption that, in general, action content is related to a high frequency of shot changes [12]. We aim at highlighting two opposite situations: video segments with a high action content (denoted "hot action") and video segments with a low action content (the opposite situation).

We tune the method parameters in order to adapt to the animated content. We have conducted an experimental test on a small set of animated movies (8 movies from CITIA [10] and Pixar Animation Company). Ten people were asked to manually browse movie contents and identify, if possible, frame segments (described as intervals $[frame_A; frame_B]$) which best fits the two generic action categories, namely: "*hot action*" (corresponding to movie segments with an intense action content, e.g. fast changes, fast motion, visual effects, etc.) and "*low action*" (mainly static scenes). For each manually labeled action segment, we compute the mean shot change ratio, $\bar{v}_T$, to capture the corresponding changing rhythm. Some of the results are presented in Table 1. Then, we compute the overall $\bar{v}_T$ mean values over all the segments within each action category, as well as the standard deviation. Having these pieces of information, we determine the intervals of $\zeta_T(i)$ values which correspond to each type of action content, as $[E\{\bar{v}_T\} - \sigma_{\bar{v}_T}; E\{\bar{v}_T\} + \sigma_{\bar{v}_T}]$. The results are synthesized with Table 2.

TABLE 1. Movie rhythm versus action content.

Movie [frames]	Segment [s]	Length	$\bar{v}_T$
"Hot action"			
François le Vaillant	2961-3443	19	3.51
	9581-10134	22	3.82
	11456-11812	14	3.25
Ferrailles	5303-5444	6	5
	8391-8657	11	3.38
Circuit Marine	7113-7401	11	3.7
The Lyon and the Song	14981-15271	12	2.33
Toy Story	2917-3582	27	2.75
	99962-101090	45	3.84
	101710-102180	19	4.43
Le Moine et le Poisson	6428-6775	14	3.5
"Low action"			
Le Trop Petit Prince	633-1574	38	0.31
	6945-8091	46	0.37
François le Vaillant	4257-6523	91	0.18
	6898-7683	31	0.38
A Bug's Life	4662-5535	35	0.17
	37209-38769	62	0.62
	66027-67481	58	0.46

Once we determine the correspondence between action perception and $\bar{v}_T$ values, we use a straightforward approach to highlight video segments which

TABLE 2. Action groundtruth.

Action type	"hot action"	"low action"
$E\{\bar{v}_T\}$	3.65	0.48
$\sigma_{\bar{v}_T}$	0.85	0.23
interval	2.8-∞	0.25-0.71

show a high number of shot changes, i.e. $\zeta_T > 2.8$ and thus candidates for "hot action" label, and a reduced number of shot changes, i.e. $\zeta_T < 0.71$ or corresponding to low action. To reduce over-segmentation of action segments, we merge neighboring action segments (within same label) at a time distance below T seconds (the size of the time window). Further, we remove unnoticeable and irrelevant action segments by erasing small action clips less than the analysis time window T. Finally, all action clips containing less than $N_s = 4$ video shots are being removed. Those segments are very likely to be the result of false detections, containing one or several gradual transitions (e.g. a "fade-out" - "fade-in" sequence).

Based on this information, action content is described with two parameters, namely the hot-action ratio (denoted HA) and the low-action ratio (denoted LA), defined thus:

$$HA = \frac{T_{HA}}{T_{total}}, \quad LA = \frac{T_{LA}}{T_{total}} \tag{1}$$

where T_{HA} and T_{LA} represent the total length of hot and low action segments, respectively, and T_{total} is the movie total length.

Gradual transition ratio. The last parameter is related to the amount of the gradual transitions used within the movie. Gradual transitions have a well defined meaning in the movie's narration. For instance a dissolve may be used to change the time of the action, similarly, a fade is used to change the action or, used in a fade group, introduces a pause before changing the action place, etc. High amounts of gradual transitions are related to a specific movie contents, for instance many artistic animated movies basically replace cuts with gradual transitions, which confers mystery to the movie (see movies "Paradise", "Cœur de Secours", "Le Moine et le Poisson", [10]). Therefore, we compute the gradual transition ratio (GT):

$$GT = \frac{T_{dissolves} + T_{fade-in} + T_{fade-out}}{T_{total}} \tag{2}$$

where T_x represents the total duration of all the gradual transitions of type x.

4. Color descriptors

One of the main particularities of the animated content is in the color distribution. Contrary to natural movies, animated movies tend to have specific color palettes, highly saturated colors, a color distribution derived from the variation of very few hues, color contrasts, very uniform color regions, etc. Therefore, we aim at capturing these properties by describing the movie's global color contents such as using statistics of color distribution (e.g. cold, warm, saturated), elementary hues, color properties and relationship of colors. This is carried out using an adaptation of the approach proposed in [7].

Prior to the analysis, several pre-processing steps are adopted. To reduce complexity, color features are computed on a summary of the initial video. Each video shot is summarized by retaining only $p = 10\%$ of its frames as a sub-sequence centered with respect to the middle of the shot (experimental tests proved that 10% is enough to preserve a good estimation of color distribution). The retained frames are down-sampled to a lower resolution (e.g. average width around 120 pixels). Finally, true color images are reduced to a more convenient color palette. We have selected the non-dithering 216 color Webmaster palette due to its consistent color wealth and the availability of a color naming system. Color mapping is performed using a minimum $L^*a^*b^*$ Euclidean distance approach applied using a Floyd-Steinberg dithering scheme. The proposed color parameters are determined as follows.

Global weighted color histogram is computed as the weighted sum of each shot color histogram:

$$h_{GW}(c) = \sum_{i=0}^{M} \left[\frac{1}{N_i} \sum_{j=0}^{N_i} h^j_{shot_i}(c) \right] \cdot \frac{T_{shot_i}}{T_{total}} \tag{3}$$

where M is the total number of video shots, N_i is the total number of the retained frames for shot i (we use temporal sub-sampling), $h^j_{shot_i}$ is the color histogram of frame j from shot i, c is a color index from the Webmaster palette (we use color reduction), and T_{shot_i} is the length of shot i. The longer the shot, the more important its contribution to the global histogram of the movie.

Elementary color histogram describes the distribution of elementary hues in the sequence:

$$h_E(c_e) = \sum_{c=0}^{215} h_{GW}(c)|_{Name(c_e) \subset Name(c)} \tag{4}$$

where c_e is an elementary color from the Webmaster color dictionary (colors are named according to color hue, saturation, and intensity), and $Name()$ returns a color's name from the palette dictionary.

Color properties. The next parameters aim at describing, first, color perception by means of light/dark, saturated/non-saturated, warm/cold color usage and second, color wealth by quantifying color variation and diversity. For instance, the light color ratio, P_{light}, reflects the percentage of bright colors in the movie:

$$P_{light} = \sum_{c=0}^{215} h_{GW}(c)|_{W_{light} \subset Name(c)} \tag{5}$$

where c is a color whose name contains one of the words defining brightness, and $W_{light} \in \{$"light", "pale", "white"$\}$. Using the same reasoning and keywords specific to each property, we define dark color ratio (P_{dark}), hard saturated color ratio (P_{hard}), weak saturated color ratio (P_{weak}), warm color ratio (P_{warm}) and cold color ratio (P_{cold}).

Additionally, we capture movie color richness with two parameters: color variation, P_{var}, which is the number of significantly different colors, and color diversity, P_{div}, defined as the number of significantly different color hues [7].

Color relationship. The final two parameters are related to the concept of perceptual relation of color in terms of adjacency and complementarity. P_{adj} reflects the amount of similar perceptual colors in the movie (neighborhood pairs of colors on a perceptual color wheel, e.g. Itten's color wheel), and P_{compl} reflects the amount of opposite perceptual color pairs (antipodal).

5. Experimental results

To test the representative power of the proposed content descriptors we use a data set consisting of 749 sequences (up to 159 hours). The animated genre is represented with 209 sequences (54 hours), namely: artistic animated movies (source [10]), films and cartoon series (source Disney, Pixar, DreamWorks animation companies). The non animated genre is represented with 541 sequences (105 hours) consisting of various genres, namely: 320 commercials (4 hours, source 1980th TV commercials and David Lynch clips; many clips include animated graphics); 74 documentaries (32 hours, both outdoor and indoor series, source BBC, IMAX, Discovery Channel); 57 movies (43 hours, both long movies and soap series, e.g. Friends, X-Files); 43 news broadcasting (19 hours, source TVR Romanian National Television Channel); 16 sports (4 hours, mainly soccer and outdoor extreme sports); 30 music clips (3 hours, source MTV Channel: dance, pop, techno music).

5.1. Setup. The classification experiments were carried out under the Weka environment [16] which provides a great perspective on the existing machine learning techniques. In our evaluations we use a large variety of supervised classification techniques, from simple Bayes to function based classification, lazy algorithms, rule based and classification trees (from each category we selected the most representative methods). Algorithm parameters were set based on preliminary experimentations.

As the choice of training data may distort the accuracy of the results, we use a cross validation approach. The data set is split into train and test sets. We use different values for the percentage split, from 10% to 90%. For a certain amount of training data, in order to shuffle all sequences, classification is repeated for all possible combinations between the train and test data. Additionally, we test different combination of descriptors. Data fusion is carried out with an early fusion approach (e.g. simple descriptor concatenation).

To assess performance we use several measures. At genre level we compute average precision (P) and recall (R) (averaged over all repetitions), which account for the number of false classification and misclassifications, respectively, thus:

$$P = \frac{\overline{TP}}{\overline{TP} + \overline{FP}}, \quad R = \frac{\overline{TP}}{\overline{TP} + \overline{FN}} \tag{6}$$

where $\overline{TP}$, $\overline{FP}$ and $\overline{FN}$ represent the *average* number of true positives, false positives and false negatives, respectively. As a global measure, we compute F_{score} and average correct classification ($\overline{CD}$), thus:

$$F_{score} = 2 \cdot \frac{P \cdot R}{P + R}, \quad \overline{CD} = \frac{\overline{N_{GD}}}{N_{total}} \tag{7}$$

where $\overline{N_{GD}}$ is the average number of good classifications and N_{total} is the number of test sequences. Each experiment is presented in the sequel.

5.2. **Classification results and discussion.** In Fig. 2 we summarize the most relevant results which were obtained for the case of fusing all temporal and color descriptors. Regarding the choice of the classification technique, some of the best results are obtained with tree based classification, which outperforms popular choices such as SVMs.

Table 3 details the classification performance for the Random Forest, which provided the highest accuracy. In this case, we obtain $F_{score} \in [72\%; 86\%]$ and average correct classification $\overline{CD} \in [85.7\%; 92.7]$, which is quite a good result considering the diversity of the test data set. With respect to false classification and misclassification, we obtain a precision up to 91.1% ($> 90\%$ for more than 60% training) compared to the recall which is only up to 80.9%. One may observe, that even using a very reduced amount of training, the results are still significant. For instance, for only 20% training, i.e. 150 sequences (from which 42 animated movies, see Table 3) and testing on 599 sequences (from which 167 animated movies), we achieve F_{score} close to 80%, i.e. only 22 false classifications and 44 misclassifications. Correct classification is in this case up to 89%, which means that from 599 sequences 533 were correctly labeled into one of the two classes.

In Fig. 3 we plot precision against recall for different descriptor combinations and amounts of training. In terms of representative power, temporal information proves to be less discriminative compared to color histograms (also due to the reduced number of features). The use of h_{GW} global histogram provides better precision than the use of h_E elementary histogram, which in turn achieve better recall. However, the best overall performance is achieved using all the descriptors even for a reduced amount of training (e.g. at least 20%, see the Blue line in Fig. 3).

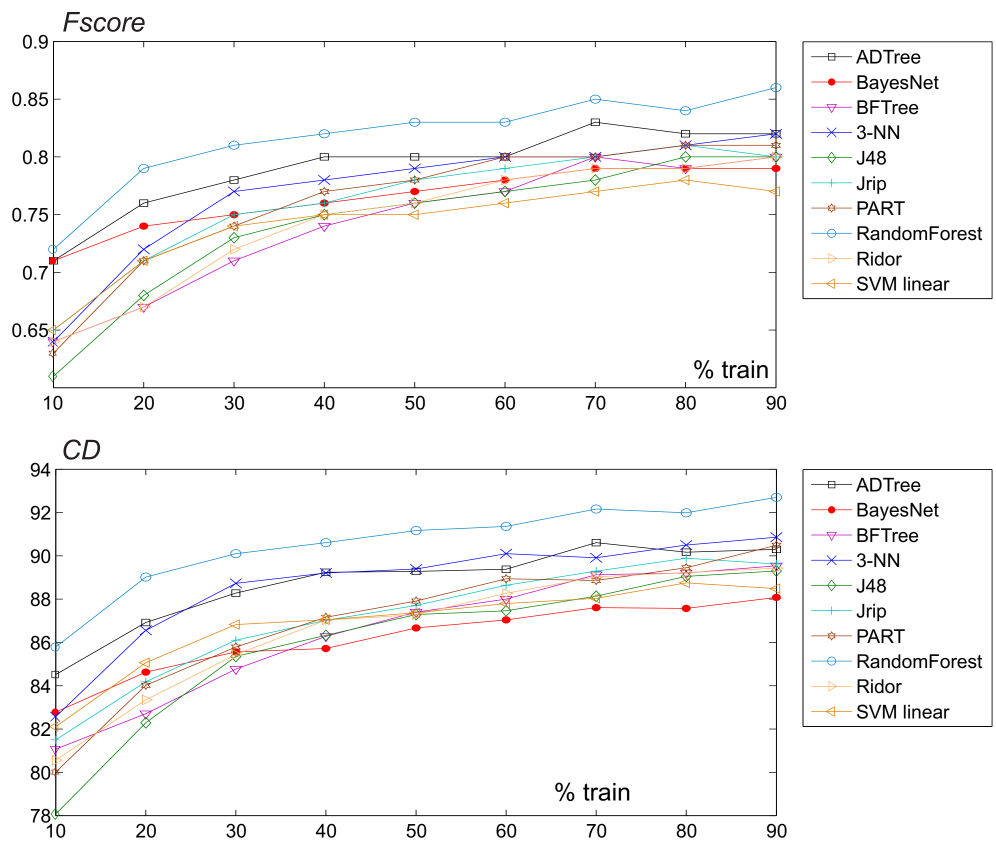

FIG. 2. Overall average F_{score} and correct classification $\overline{CD}$ (see eq. 7) for various machine learning techniques and different amounts of training data.

TABLE 3. Random Forest on all action and color descriptors.

train (%)	# train seq.	# train anim.	# test seq.	# test anim.	P (%)	R (%)	$\overline{TP}$	$\overline{FP}$	$\overline{FN}$	F_{score} (%)	$\overline{CD}$ (%)
90	675	189	74	20	**91.1**	**80.9**	16.2	1.6	3.8	85.7	92.7
80	600	168	149	41	**90.9**	**78.7**	32.3	3.2	8.7	84.4	92
70	525	147	224	62	**90.8**	**79.8**	49.5	5	12.5	84.9	92.2
60	450	126	299	83	**89.8**	**77.7**	64.5	7.4	18.5	83.3	91.4
50	375	105	374	104	**89.1**	**77.8**	80.9	9.9	23.1	83.1	91.2
40	300	84	449	125	**88**	**76.7**	95.9	13	29.1	82	90.6
30	225	63	524	146	**87**	**75.8**	110.6	16.5	35.4	81	90.1
20	*150*	*42*	*599*	*167*	***84.9***	***73.8***	*123.2*	*21.9*	*43.8*	*79*	*89*
10	75	21	674	188	**79.5**	**66.1**	124.3	32.1	63.7	72.2	85.8

6. Conclusions and future work

We addressed a particular case of video genre classification, i.e. the classification of the animated genre. We proposed two categories of content descriptors which are adapted to animated contents, namely: *temporal descriptors*

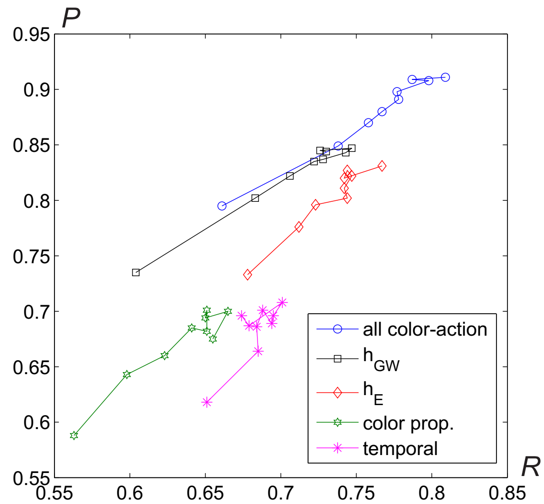

FIG. 3. Average precision vs. recall (see eq. 6) obtained with Random Forest and various descriptor combinations (the amount of training data increases along the curves).

and *color descriptors*. These descriptors were used with several binary classification schemes to classify video footage into animated and non animated content.

To provide a pertinent evaluation tests were performed on an extensive data set, namely 749 sequences containing various genres of animated movies, but also other video genres: commercials, documentaries, movies, news, sport and music.

We achieve very promising results using the combination of all descriptors, e.g. an average global correct detection ratio and F_{score} up to 92.7% and 85.7%, respectively. Future work should push forward descriptors to a higher semantic level, such as exploiting human concepts.

Acknowledgments

This work has been co-funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557.

The authors would like to thank CITIA - The City of Moving Images and Folimage Animation Company for the animated movies.

REFERENCES

[1] B. Ionescu, C. Vertan, P. Lambert, A. Benoit, "A Color-Action Perceptual Approach to the Classification of Animated Movies", ACM International Conference on Multimedia Retrieval, Trento, Italy, 17-20 April, 2011.

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

[3] T.I. Ianeva, A.P. Vries and H. Rohrig, "Detecting cartoons: a case study in automatic video-genre classification", IEEE International Conference on Multimedia and Expo, 1, pp. 449-452, 2003.

[4] M. Roach, J. S. Mason, and M. Pawlewski, "Motion-based classication of cartoons", International Symposium on Intelligent Multimedia, pp. 146-149, 2001.

[5] R. Glasberg, A. Samour, K. Elazouzi and T. Sikora, "Cartoon-recognition using video and audio-descriptors", 13th European signal processing conference, Antalya, Turkey, 2005.

[6] X. Gao, J. Li and N. Zhang, "A Cartoon Video Detection Method Based on Active Relevance Feedback and SVM", Springer Lecture Notes in Computer Science, 3972, pp. 436-441, 2006.

[7] B. Ionescu, D. Coquin, P. Lambert and V. Buzuloiu: "A Fuzzy Color-Based Approach for Understanding Animated Movies Content in the Indexing Task", Eurasip Journal on Image and Video Processing, doi:10.1155/2008/849625, 2008.

[8] D. Brezeale, D.J. Cook, "Automatic Video Classification: A Survey of the Literature", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 38(3), pp. 416-430, 2008.

[9] M. Montagnuolo, A. Messina, "Parallel Neural Networks for Multimodal Video Genre Classification", Multimedia Tools and Applications, 41(1), pp. 125-159, 2009.

[10] CITIA - Animaquid Animated Movie Indexing System, http://www.annecy.org/home/index.php? Page_ID=44.

[11] B. Ionescu, V. Buzuloiu, P. Lambert, D. Coquin, "Improved Cut Detection for the Segmentation of Animation Movies", IEEE International Conference on Acoustic, Speech and Signal Processing, Toulouse, France, 2006.

[12] H.W. Chen, J.-H. Kuo, W.-T. Chu and J.-L. Wu, "Action movies segmentation and summarization based on tempo analysis", ACM International Workshop on Multimedia Information Retrieval, pp. 251- 258, New York, 2004.

[13] W.A.C. Fernando, C.N. Canagarajah, D.R. Bull, "Fade and Dissolve Detection in Uncompressed and Compressed Video Sequence", IEEE International Conference on Image Processing, Kobe, Japan, pp. 299-303, 1999.

[14] B. Ionescu, C. Vertan, P. Lambert, "Dissolve Detection in Abstract Video Contents", IEEE International Conference on Acoustic, Speech and Signal Processing, Prague, Czech Republic, 22-27 May, 2011.

[15] I. H. Witten, E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques", Second Edition, Eds. Morgan Kaufmann, ISBN 0-12-088407-0, 2005.

[16] Weka Data Mining with Open Source Machine Learning Software in Java, University of Waikato, `http://www.cs.waikato.ac.nz/ml/weka/`, 2011.