Automatic Categorization of Shots in News Videos Based on the Temporal Relations

Kazimierz Choroś(✉)

Department of Information Systems, Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
kazimierz.choros@pwr.edu.pl

Abstract. The development of new methods and technologies of video indexing and retrieval is stimulated by the growing amount of digital video data stored in Internet video collections, TV shows archives, video-on-demand systems, personal video archives offered by Web services, etc. The videos very frequently offered in the Web by broadcast channels are news videos and sports news videos. Content-based indexing of videos is based on the automatic detection of a video structure. A video shot is the main structural video unit. Shots can be of different categories such as intro or final animation, chart or table shots, anchor, reporter, statement, or interview shots, and finally the most informative report shots. The temporal aggregation results in grouping of shots into scenes of a given category. The paper examines the usefulness of the temporal aggregation method to select report shots and non-report shots in a news video.

Keywords: Content-based video indexing · News videos · Video structures · Temporal aggregation · News shots categories · Digital video segmentation

1 Introduction

The great variety of approaches and methods of content-based video indexing are applied to an automatic processing of television broadcast. In television broadcast archives there is a huge number of news and TV sports news videos from last years as well as from the past. Many old analogue videos are digitized, new news videos are being stored every day. Very efficient methods of automatic content-based analyses of digital videos are strongly desirable. The main goal of content-based video indexing of broadcast news videos is to ensure an effective retrieval of special events or special people, of official statements or political polemics and commentaries, etc. Whereas, in the case of sports news the main purpose is to detect players and games, or to select reports of a given sports category. To achieve this purpose, the automatic categorization of sports events, i.e. the automatic detection of the sports disciplines of reported events should be provided. In consequence, the effective retrieval of sports news and of sports highlights of a given sports category such as the best or actual games, tournaments, matches, contests, races, cups, etc., special player behaviours or actions like penalties, jumps, or race finishes, etc. becomes possible.

M. Núñez et al. (Eds.): ICCCI 2015, Part II, LNCS 9330, pp. 13–23, 2015.
DOI: 10.1007/978-3-319-24306-1_2

In most approaches the content-based indexing method needs to analyze all frames of a video. Because such a procedure is extremely time consuming the indexing should be limited to the key-frames, to only one or to a few frames for every shot or for every video scene. The detection of a key-frame of a shot or of a scene requires very effective temporal segmentation methods.

The automatic detection of transitions between shots in a digital video with the purpose of temporal segmentation is relatively well managed and applied in practice. Unfortunately, the detection of scenes is not effectively carried out. A scene is usually defined as a group of consecutive shots sharing similar visual properties and having a semantic correlation – following the classical rule of unity of time, place, and action. The temporal aggregation method [1] detects player scenes taking into account only shot lengths. A player scene is a scene presenting the sports game, i.e. a given scene was recorded on the sports fields such as playgrounds, tennis courts, sports hall, swimming polls, ski jumps, etc. All other non-player shots and scenes usually recorded in a TV studio such as commentaries, interviews, charts, tables, announcements of future games, discussions of decisions of sports associations, etc. are called studio shots or studio scenes. Studio shots are not useful for video categorization and therefore should be rejected. It was observed that the studio scenes may be even two thirds of sports news. This rejection of non-player scenes before starting content analyses creates a great opportunity to significantly reduce computing time and conduct these analyses more efficiently.

Generally different video genres have different editing style. The specific nature of videos has an important influence on the efficiency of temporal segmentation methods. In the experiments performed in [2] the efficiency of segmentation methods was analyzed for five different categories of movies: TV talk-show, documentary movie, animal video, action & adventure, and pop music video. It has been shown that the segmentation parameters should be suitable to the specificity of the videos.

TV news video editing is similar to that of TV sports news but shots are longer in average. Then the statements and commentaries can be more significant in news than in sports news because these statements are not spoken by anchorman but also for example by politicians or famous people. The detection of politicians is important and may be realized using for example face detection methods.

In this paper the usefulness of the temporal aggregation in detection of report and non-report shots in a news video is verified. The paper is organized as follows. The next section describes related work in the area of an automatic news and sports news shot categorization. The main idea of the temporal aggregation method is presented and the detection of pseudo-scenes using temporal aggregation is outlined in the third section. The forth section presents the experimental results of the detection of the main structural units of TV news obtained in the AVI Indexer. The final conclusions and the future research work areas are discussed in the last fifth section.

2 Related Work

Much research has been carried out in the area of automatic recognition of video content and of visual information indexing and retrieval [3–7]. Traditional textual techniques frequently applied for videos are not sufficient for nowadays video archive browsers. The effective methods of the automatic categorization of a huge amount of broadcast news videos – mainly sports news videos – would be highly desirable. Most of proposed methods require the detection of the structure of videos being indexed and the categorization of shots and scenes detected in videos [8].

Different criteria can be used for the shot categorization in indexing process, the most interesting criterion is the content of a shot. In news videos such shot categories can be defined as [9]: anchor shots, animation (intro), communication (static image of reporter), interview, reporter, maps, studio (discussion with a guest), synthetic (tables, charts, diagrams), whether, and of course report shots.

Other authors developed methods for sports shots classification to such classes as: court views, players and coach, player close-up views, audience views, and setting long views [10], but also as: long views, middle views, close-up views, out of fields views [11], as well as like: intro, headlines, player shots, studio shots [12]. In [13] all sports shots from all types of field video are classified basing on the perceived distance between the camera and the object presented in the shot. Fourteen different shot classes were defined: close up shot head with simple background, close up shot head complex background, close up shot head mixture background, close up shot waist up simple background, close up shot waist up complex background, close up shot waist up mixture background, short distance shot presenting player(s) simple background, short distance shot presenting player(s) complex background, short distance shot presenting player(s) mixture background, short distance shot presenting spectators, long distance shot presenting centre of the field, long distance shot presenting right side of the field, long distance shot presenting left side of the field, and long distance shot presenting spectators.

Anchor/non-anchor shots are frequently used as a starting point for the automatic recognition of a news or sports news video structure. Anchorperson shot detection is still a challenging and important stage of news video analysis and indexing. Recent years, many algorithms have been proposed to detect anchorperson shots. Because we observe the very high similarity between anchor shots (very static sequences of frames, small changes, the same repeated background) one of the approaches of an anchor shot detection is based on template matching. Whereas the other methods are based on different specific properties of anchor shots or on temporal analyses of shots. In the first group of methods a set of predefined models of an anchor should be defined and then, they are matched against all frames in a news video, in order to detect potential anchor shots. The second group of an anchor shot detection methods is mainly based on clustering. Unfortunately, the proposed methods are very time-consuming because they require complex analyses of a great number of video frames. The third approach based on temporal aggregation is very fast because only shot durations are analysed.

The high values of recall and precision for anchorperson detection have been obtained in the experiments on 10 news videos [14]. The news videos were firstly as usually segmented into shots by a four-threshold method. Then the key frames were extracted from each shot. The anchorperson detection was conducted from these key frames by using a clustering-based method based on a statistical distance of Pearson's correlation coefficient.

The new method presented in [15] can be also used for dynamic studio background and multiple anchorpersons. It is based on spatio-temporal slice analysis. This method proposes to extract two different diagonal spatio-temporal slices and divide them into three portions. Then all slices from two sliding windows obtained from each shot are classified to get the candidate anchor shots. And finally, the real anchor shots are detected using structure tensor. The experiments carried out on news programs of seven different styles confirmed the effectiveness of this method.

The algorithm described in [16] analyzes audio, frame and face information to identify the content. These three elements are independently processed during the cluster analysis and then jointly in a compositional mining phase. The temporal features of the anchorpersons for finding the speaking person that appears most often in the same scene are used to differentiate the role played by the detected people in the video. Significant values of precision and recall have been obtained in the experiments carried out for broadcast news coming from eight different TV channels.

A novel anchor shot detection method proposed in [17] detects an anchorperson cost-effectively by reducing the search space. It is achieved by using skin colour and face detectors, as well as support vector data descriptions with non-negative matrix factorization.

It is observed that the most frequent speaker is the anchorman [18]. An anchor speaks many times during the programme, so the anchorperson shots are distributed all along the programme timeline. This observation leads to the selection of the speaker who most likely is the anchorman. It is assumed that a speaker clustering process labels all the speakers present in the video and associates them to temporal segments of the content. However, there are some obvious drawbacks, because a shot with a reporter (interview shots) or with a politician (statement shots) frequently found in news can be erroneously recognized as an anchor shot.

Another observation in a large database [19] draws much attention to interview scenes. In many interview scenes an interviewer and an interviewee recursively appear. A technique called interview clustering method based on face similarity can be applied to merge these interview units.

In [20] a fast method of automatic detection of anchorperson shots has been presented. The method is useful for detection long duration shots such as anchor, reporter, interview, or any other statement shots.

Video analyses discussed in the related papers as well as in this research are the methods using visual features only. There are also audio-visual approaches analyzing not only visual information but also audio (see for example [21, 22]).

3 Temporal Segmentation and Aggregation in the AVI Indexer

The Automatic Video Indexer AVI [23] is a research system designed to develop new tools and techniques of automatic video content-based indexing for retrieval systems, mainly based on the video structure analyses [24] and using the temporal aggregation method [1]. The standard process of automatic content-based analysis and video indexing is composed of several stages. Usually it starts with a temporal segmentation resulting in the segmentation of a movie into small units called video shots. Shots can be grouped to make scenes, and then key-frame or key-frames for every scene can be selected for further analyses. In the case of TV sports news every scene is categorized using such strategies as: detection of playing fields, of superimposed text like player or team names, identification of player faces, detection of lines typical for a given playing field and for a given sports discipline, recognition of player and audience emotions, and also detection of sports objects specific for a given sports category. Whereas in the case of TV news scenes can be categorized basing on the people or place detection using face detection or object detection.

The detection of video scenes facilities the optimization of indexing process. The automatic categorization of news videos will be less time consuming if the analyzed video material is limited only to scenes the most adequate for content-based analyses like player scenes in TV sports news or official statements in TV news. The temporal aggregation method implemented in the AVI Indexer is applied for a video structure detection. The method detects and aggregates long anchorman shots. The shots are grouped into scenes basing on the length of the shots as a sufficient sole criterion.

The temporal aggregation method has two main advantages. First of all it detects player scenes, therefore the most informative parts of sports news videos. Then, it significantly reduces video material analyzed in content-based indexing of TV sports news because it permits to limit indexing process only to player scenes. Globally, the length of all player scenes is significantly lower than the length of all studio shots.

The temporal aggregation is specified by three values: minimum shot length as well as lower and upper limits representing the length range for the most informative shots. The values of these parameters should be determined taking into account specific editing style of a video and its high-level structure.

Formally, the temporal aggregation process is defined as follows [25]:

- single frame detected as a shot is aggregated to the next shot,
  if $(L(shot_i) == 1\ [frame])$ *then* $L(shot_{i+1}) = L(shot_{i+1}) + 1;$
  $LS = LS - 1;$
  where $L(shot_i)$ *is the length [measured in frames] of the detected shot i and* $shot_{i+1}$ *is a next shot on a timeline and LS is a number of shots detected;*
- very short shots should be aggregated till their aggregated length attains a certain value Min_Shot_Length,
  while $((L(shot_i) < MIN\_Shot\_Length)$ *and* $(L(shot_{i+1}) < MIN\_Shot\_Length))$ *do*
  $\{ L(shot_i) = L(shot_i) + L(shot_{i+1});$
  $LS = LS - 1; \}$

- all long consecutive shots should be aggregated because these shots seem to be useless in further content analyses and categorization of sports events,
  while ((L(shot$_i$) > MAX_Shot_Length) and (L(shot$_{i+1}$) > MAX_Shot_Length)) do
  { L(shot$_i$) = L(shot$_i$) + L(shot$_{i+1}$);
  LS = LS – 1; }
- after aggregation all shots of the length between two a priori defined maximum and minimum values should remain unchanged – these shots are very probably the most informative shots for further content-based analyses.

Very short shots including single frames are relatively very frequent. Generally very short shots of one or several frames are detected in case of dissolve effects or they are simply wrong detections. The causes of false detections may be different [26]. Most frequently it is due to very dynamic movements of players during the game, very dynamic movements of objects just in front of a camera, changes (lights, content) in advertising banners near the player fields, very dynamic movements of a camera during the game, light flashes during games or interviews. These extremely short shots resulting from temporal segmentation are joined with the next shot in a video. So, the first two steps of the temporal aggregation of shots also leads to the significant reduction of false cuts incorrectly detected during temporal segmentation.

4 Report and Non-report Shots in Temporally Aggregated News

The method of temporal aggregation has been applied in the experiments performed in the AVI Indexer. The temporal aggregation has been used with such parameters that only shots of the duration not lower than 45 frames (MIN_Shot_Length) and not greater than 305 frames (MAX_Shot_Length) have been not aggregated. These are shots of the length from 2 to 12 seconds ± 5 frames of tolerance.

Six editions of the TV News “Teleexpress” used in the experiments have been broadcasted in the first national Polish TV channel (TVP1). Their characteristics before and after temporal aggregation are presented in Table 1. The “Teleexpress” is broadcasted every day and is of 15 minutes. This TV program is mainly dedicated to young people. It is dynamically edited, it is very fast paced with very quickly uttered anchor comments. So, the dynamics of the “Teleexpress” News can be comparable to the dynamics of players scenes in TV sports news. However the number of topics and events reported in the news is usually much greater than in the sports news. The question is how the temporal aggregation method changes the temporal relations of shots, whether despite the fact that shots are aggregated their temporal characteristics enable us to predict a shot category.

Table 1. Characteristics before and after temporal aggregation of the six „Teleexpress" News videos broadcasted in March 2014 (03-03, 03-05, 03-06, 03-08, 03-09, and 03-11).

	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6	Average
Length [sec.]	907	900	899	894	893	905	900
After temporal segmentation							
Total number of shots before aggregation	520	461	492	465	523	453	486
Number of shots of less than 15 frames	310	260	277	274	304	260	281
Number of shots of a single frame	261	201	222	205	223	181	216
Real number of anchor shots	13	12	13	14	13	13	13
After temporal aggregation							
Total number of shots after aggregation	176	178	173	159	197	169	175
Number of shots tipped to be reports (45<=length<=305)	161	164	157	144	182	153	160
Percentage of frames in the report shots in the video [%]	73	69	72	66	73	70	71
Number of long aggregated shots (>305 frames)	14	14	16	15	14	15	15
Number of incorrectly (two or more shots of different categories) aggregated shots	2	6	0	4	3	2	2.8

The aggregation is incorrect if the length of an aggregated shot becomes not adequate for its category, i.e. a report aggregated shot becomes so long that it can be treated as for example an anchor shot. The aggregation is also incorrect if two or more shots of different categories are aggregated, mainly if a report shot is aggregated with no report shot. It happens very rarely. Only 2.8 incorrectly aggregates shots in average have been observed (last row in Table 1). The most frequently it was a case of the aggregation of very short report shot with a subsequent anchor shot. Such a long aggregated shot would be then processed as an anchor shot, and thus it should not distort the results of content-based indexing.

The most important result is that after the application of the temporal aggregation the anchor shots are still the longest shots in news (Table 2). Although, it should be noticed that the statement, reporter, or interview shots are also at the beginning of the

ranking of longest shots in news videos. They are almost as frequent (7 shots) in long shots as report shots (8 shots). So, the shot aggregation facilitates the detection of speaking person shots. Thus, the temporal analyses of aggregated shot makes it easy to select report shots and non-report shots. And this is a key problem in indexing.

Table 2. Analysis of long aggregated shots.

	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6	Average
Number of aggregated long shots	14	14	16	15	14	15	14.67
Real number of all anchor shots	13	12	13	14	13	13	13.00
Number of anchor shots in long shots	12	11	11	13	12	11	11.67
Percentage of detected anchor shots [%]	92.31	91.67	84.62	92.86	92.31	84.62	89.77
Number of report shots	1	0	3	2	1	1	1.33
Percentage of report shots in long shots [%]	7.14	0	18.75	13.33	7.14	6.67	8.84

Between all 88 long aggregated shots there are 70 anchor shots and 7 other speaking people shots, eight report shots, one chart shot, and one final animation. The report shots represent only about 9 % of all aggregated long shots. To detect faster anchor shots it is desirable to reduce the video space by using temporal aggregation.

The structure of a news video is as follows: it starts with the intro animation, then several stories are presented and commented containing an anchor shot or shots followed by a sequence of report shots optionally enhanced by reporter, politician's statement, interview, or chart, table, diagram shots. Similarly to the began the news is finished with a final animation shot. The length of the intro and final animation is not constant, so after temporal segmentation at the beginning as well as at the end of a news video we can receive different sequence of shots. It is due to the fact that sometimes there is a fade from black to the intro part and the final animation usually fades away to a black frame. But we very often observed that the first frame of the intro or the last frame of the final animation are frozen, so the length of these parts of a news video are of an unpredictable duration.

The analyses of the lengths of both the 50 longest shots (Table 3) as well as of all shots (Table 4) in the tested videos clearly confirm that anchor shots are the longest parts of videos before and also after aggregation: 414 frames in average for the most longest shots in a news video and 341 frames in average for all shots. And at the same time report shots are the shortest parts of videos: 173 frames in average for the most longest shots in a news video and 92 frames in average for all shots.

Table 3. Number of shots of a given category and their average lengths among the 50 longest aggregated shots in each of the videos.

Shot categories	Video 1		Video 2		Video 3		Video 4		Video 5		Video 6		Averages for all videos	
Intro	1	140	1	281	1	142	1	140	1	147	0	–	1	170
Anchor shots	11	385	10	441	13	381	11	447	10	409	11	420	11	414
Chart shots									2	272	1	225	1.5	249
Statement, Reporter, or Interview	7	220	6	218	10	249	4	225	5	191	11	230	7	222
Report	28	187	27	153	25	192	30	181	29	147	24	178	27	173
Final Animation	1	300	1	306	1	246	1	282	1	326	1	330	1	298

Table 4. Number of all shots of a given category and their average lengths among in the tested videos.

Shot categories	Video 1		Video 2		Video 3		Video 4		Video 5		Video 6		Averages for all videos	
Intro	3	101	1	281	3	93	2	125	2	145	2	92	2.2	140
Anchor shots	11	385	13	292	16	345	13	385	14	310	15	329	13.7	341
Chart shots									2	272	1	225	1.5	249
Statement, Reporter, or Interview	7	220	6	218	11	236	5	203	7	163	12	218	8.0	210
Report	151	97	148	89	139	92	134	98	168	86	134	92	146	92
Final Animation	1	300	1	306	1	246	1	282	2	170	1	330	1.0	298

The most important is that the number of shots decreased almost three times. If the indexing process is based on the analysis of a single key frame for every shot in a video the processing time of indexing will be also decreased three times. This is a great advantage of the temporal aggregation.

5 Final Remarks

The methods of content-based video indexing are still being improved, new methods are still being proposed. Many of them adapted to the content analysis of news videos are based on video structure. The detection of news video structure and categorization of news video shots is very important. The key problem is to select report shots and non-report shots because usually different indexing strategies should be applied.

The temporal aggregation can successfully reduce the video space analyzed in content-base indexing without disturbing news shot categorization. Furthermore, the temporal aggregation can be also used to select report shots and non-report shots in news videos. The results of tests performed in the AVI Indexer have confirmed that the temporal aggregation facilitates the automatic parsing of video structure of news videos.

All candidate shots for non-report shots such as anchor shots as well as statement, reporter, or interview shots can be then analyzed using usually proposed approaches mainly based on face detection and person recognition. Whereas, the report shots demand a variety of methods based on different approaches.

References

1. Choroś, K.: Temporal aggregation of video shots in TV sports news for detection and categorization of player scenes. In: Bădică, C., Nguyen, N.T., Brezovan, M. (eds.) ICCCI 2013. LNCS, vol. 8083, pp. 487–497. Springer, Heidelberg (2013)
2. Choroś K., Gonet M.: Effectiveness of video segmentation techniques for different categories of videos. In: New Trends in Multimedia and Network Information Systems, pp. 34–45. IOS Press, Amsterdam (2008)
3. Truong, B.T., Venkatesh, S.: Video Abstraction: A Systematic Review and Classification. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) **3**(1), 1–37 (2007)
4. Money, A.G., Agius, H.: Video summarisation: a conceptual framework and survey of the state of the art. J. of Visual Communication and Image Representation **19**, 121–143 (2008)
5. Hu, W., Xie, N., Li, L., Zeng, X., Maybank, S.: A survey on visual content-based video indexing and retrieval. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **41**(6), 797–819 (2011)
6. Del Fabro, M., Böszörmenyi, L.: State-of-the-art and future challenges in video scene detection: a survey. Multimedia Systems **19**(5), 427–454 (2013)
7. Asghar, M.N., Hussain, F., Manton, R.: Video indexing: a survey. Int. J. of Computer and Information Technology **3**(1), 148–169 (2014)
8. Kompatsiaris, Y., Mérialdo, B., Lian, S. (eds.): TV Content Analysis: Techniques and Applications. CRC Press, Boca Raton (2012)
9. Valdés, V., Martínez, J.M.: On-line video abstract generation of multimedia news. Multimedia Tools and Applications **59**(3), 795–832 (2012)
10. Duan, L.Y., Xu, M., Tian, Q., Xu, C.S., Jin, J.S.: A unified framework for semantic shot classification in sports video. IEEE Transactions on Multimedia **7**(6), 1066–1083 (2005)
11. Lang, C., Xu, D., Jiang, Y.: Shot type classification in sports video based on visual attention. In: Proc. of Int. Conf. on Computational Intelligence and Natural Computing (CINC 2009), vol. 1, pp. 336–339. IEEE (2009)
12. Choroś, K.: Automatic detection of headlines in temporally aggregated TV sports news videos. In: Proc. of the 8th Int. Symp. on Image and Signal Processing and Analysis (ISPA 2013). IEEE, pp. 147–152 (2013)
13. Kapela, R., McGuinness, K., O'Connor, N.E.: Real-time field sports scene classification using colour and frequency space decompositions. Journal of Real-Time Image Processing, 1–13 (2014)

14. Ji, P., Cao, L., Zhang, X., Zhang, L., Wu, W.: News videos anchor person detection by shot clustering. Neurocomputing **123**, 86–99 (2014)
15. Zheng, F., Li, S., Wu, H., Feng, J.: Anchor shot detection with diverse style backgrounds based on spatial-temporal slice analysis. In: Boll, S., Tian, Q., Zhang, L., Zhang, Z., Chen, Y.-P.P. (eds.) MMM 2010. LNCS, vol. 5916, pp. 676–682. Springer, Heidelberg (2010)
16. Broilo, M., Basso, A., De Natale, F.G.: Unsupervised anchorpersons differentiation in news video. In: Proc. of the 9th International Workshop on Content-Based Multimedia Indexing (CBMI), pp. 115–120. IEEE (2011)
17. Lee, H., Yu, J., Im, Y., Gil, J.M., Park, D.: A unified scheme of shot boundary detection and anchor shot detection in news video story parsing. Multimedia Tools and Applications **51**(3), 1127–1145 (2011)
18. Montagnuolo, M., Messina, A., Borgotallo, R.: Automatic segmentation, aggregation and indexing of multimodal news information from television and the Internet. Int. J. of Information Studies **1**(3), 200–211 (2010)
19. Dong, Y., Qin, G., Xiao, G., Lian, S., Chang, X.: Advanced news video parsing via visual characteristics of anchorperson scenes. Telecommunication Systems **54**(3), 247–263 (2013)
20. Choroś, K.: Automatic fast detection of anchorperson shots in temporally aggregated TV news videos. In: Nguyen, N.T., Trawiński, B., Kosala, R. (eds.) ACIIDS 2015. LNCS, vol. 9012, pp. 339–348. Springer, Heidelberg (2015)
21. El Khoury, E., Sénac, C., Joly, P.: Audiovisual diarization of people in video content. Multimedia Tools and Applications **68**(3), 747–775 (2014)
22. Qu, B., Vallet, F., Carrive, J., Gravier, G.: Content-based inference of hierarchical structural grammar for recurrent TV programs using multiple sequence alignment. In: Proc. of the IEEE International Conference on Multimedia and Expo., ICME, pp. 1–6 (2014)
23. Choroś, K.: Video structure analysis and content-based indexing in the automatic video indexer AVI. In: Nguyen, N.T., Zgrzywa, A., Czyżewski, A. (eds.) Advances in Multimedia and Network Information System Technologies. AISC, vol. 80, pp. 79–90. Springer, Heidelberg (2010)
24. Choroś, K.: Video structure analysis for content-based indexing and categorisation of TV sports news. Int. J. of Intelligent Information and Database Systems **6**(5), 451–465 (2012)
25. Choroś, K.: Automatic detection of headlines in temporally aggregated TV sports news videos. In: Proc. of the 8th International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 147–152. IEEE (2013)
26. Choroś, K.: False and miss detections in temporal segmentation of TV sports news videos – causes and remedies. In: Zgrzywa, A., Choroś, K., Siemiński, A. (eds.) New Research in Multimedia and Internet Systems. Advances in Intelligent Systems and Computing 314, pp. 35–46. Springer, Heidelberg (2014)

Springer

http://www.springer.com/978-3-319-24305-4

Computational Collective Intelligence
7th International Conference, ICCCI 2015, Madrid, Spain, September 21-23, 2015, Proceedings, Part II
Núñez, M.; Nguyen, N.T.; Camacho, D.; Trawiński, B. (Eds.)
2015, XXII, 661 p. 220 illus., Softcover
ISBN: 978-3-319-24305-4