http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 9 (5): 1003-1012, 2010
ISSN 1812-5638

Video Macrosegmentation Using Automatic Analysis of Similarity Matrices

S. Haidar, B. Chebaro and B. Haidar
Department of Computer Sciences, Faculty of Sciences I, Lebanese University, Lebanon

Abstract: In the present study, we propose an automatic method for segmenting video material using the comparison matrix. Without a priori information, neither any training phase, the similarity matrices are filtered then regional minimums and maximums are extracted over the line and column projections. Pseudo-homogeneous areas are identified between each pair of regional minimums. The pseudo-homogeneous areas in the resulting matrix are thus the macro-segment video records in question. We experiment our method over a database of short video commercials to illustrate the potential of our approach. An overall evaluation is finally established and given.

Key words: Semantic video analysis, segmentation, comparison, audiovisual features

INTRODUCTION

The uses of video documents are related to the narration, to their different structure and also the diversity of forms and contents of these documents: news, documentaries, films, magazines, talk shows, sports broadcasts, emissions variety, music videos, etc and their styles particularly in fiction documents. The rapid evolution of narrative construction, editing and postproduction techniques has profoundly changed these practices. Video documents comparison using similarity matrix is one of the methods proposed in the field of semantic analysis. Many propose to realize the comparison at the level of macro-segments.

The macro-segment was originally proposed by Joly and Aigrain (1996). This term has been defined as a segment corresponding to more than one plan in the document but still corresponding to less than the entire document. The concept of macro-segment is more semantic and therefore make difficult or impossible, any possible general and descriptive definition to guide macro-segmentation algorithms.

The main applications of macro-segmentation, through the implementation of a hierarchical structuring of audiovisual media in which it participates, include: navigation and non-linear access to content, the representation of documents (Yeung *et al*., 1998), the creation of summaries of videos (or trailer) (Peyrard and Bouthemy, 2002), the generation of indexes or tables of contents (Rui *et al*., 1999), searching for significant events, the classification of sequences according to their kind, or query by content at different levels of granularity.

It is not easy to offer an overview of macrosegmentation methods because of the diversity of the concept of macrosegment. Different reading schemes have been proposed, some differentiate methods related to compression, other distinct methods called semantic and those called syntax (Gunsel *et al*., 1998a).

Within the hierarchical structure, the main approaches supporting the methods implemented are:

- Methods based on shots grouping based on both physical and temporal similarity (Yeung *et al*., 1998; Galmar *et al*., 2008)
- Methods based on using a priori information (Carrive *et al*., 2000; Zhang *et al*., 1994; Ariki and Saito, 1996; Gunsel *et al*., 1998a, b; Merlino *et al*., 1997; Joly and Aigrain, 1996; Aigrain *et al*., 1997; Feng *et al*., 2008). New rules were proposed and the macrosegmentation is based on these rules and requires the extraction of primitives for streaming audio and video and formalizations of editing techniques (Joly and Aigrain, 1996). The rules, derived from a thorough study of audiovisual materials, theory of films and discussions with professionals were finally defined in Aigrain *et al*. (1997). On the other hand, the methods related to a stratification approach resulted in obtaining a temporal sequence level of abstraction not unlike that of macro-segments (Gunsel *et al*., 1998a)
- Methods based on joint use of different types of information present in the audiovisual document (Saraceno and Leonardi, 1997; Lienhart *et al*., 1999; Adams *et al*., 2000)

Corresponding Author: Siba Haidar, Department of Computer Sciences, Faculty of Sciences I, Lebanese University, Lebanon
Tel: 009613940694 Fax: 009615460696

SIMILARITY MATRIX

Many researchers were interested in searching in large databases (Sivaselvan and Gopalan, 2006), or in comparison between objects. Some have suggested the similarity matrix as a mean to solve their problems. The similarity matrices were featured in various fields, whether the field of statistics (Sickle, 1997), ecology (Daget and Durand, 1968), bioinformatics (Collins and Coulson, 1990), or in the field of multimedia (Yahya and Abdalla, 2003).

For example, in the field of multimedia:

- Cooper and Foote (2001) detected scene boundaries by considering the self-similarity across time and calculated for each moment in the video, the self-similarity for past and future regions and for cross-similarity, that is to say, between past and future of a region
- In earlier works, we proposed and defined the similarity matrices for the comparison of audiovisual documents (Haidar *et al*., 2004-2006). We give in the next paragraph a short recall summarizing the main ideas of the method

Similarity matrix of Haidar *et al.*: To assess the similarity between video materials, we look for the common elements. Each video is represented by a set of its audiovisual features (dominant color, motion, contrast, etc.). These characteristics when extracted for a video document consisting of 25 frames per second, form time series. The role of the similarity matrix is to look for similar sequences between the two sets of series. For example, given a characteristic Fx, time series representing this characteristic are extracted from both compared documents. Over the two series that we obtain, we apply our multidimensional sequence matching algorithm and hence a square matrix is created (Fig. 1) whose axes are proportional to the temporal dimension of each sequence and whose unit step is tMin.

In fact, our sequence matching method, like most of the matching methods, in an environment with a large set of data sequences, works in two phases:

- In the first phase, only a finite number of data sequences are kept after a filtering process. We consider that these sequences are matching candidates
- In the second phase, all candidate sequences are verified for the actual matching using a morphological filter

Once potential candidates are filtered, given two sequences, we shall proceed to the resemblance verification. To achieve this, we construct a one-dimensional morphological envelope using erosion and dilation. The matched sequences have a length in time t which may vary between tMin and tMax boundaries. The algorithm works on deep down comparison. It continues to compare until it finds all possible similar segments of different lengths. Finally, we obtain variable length couples judged as similar for the considered feature.

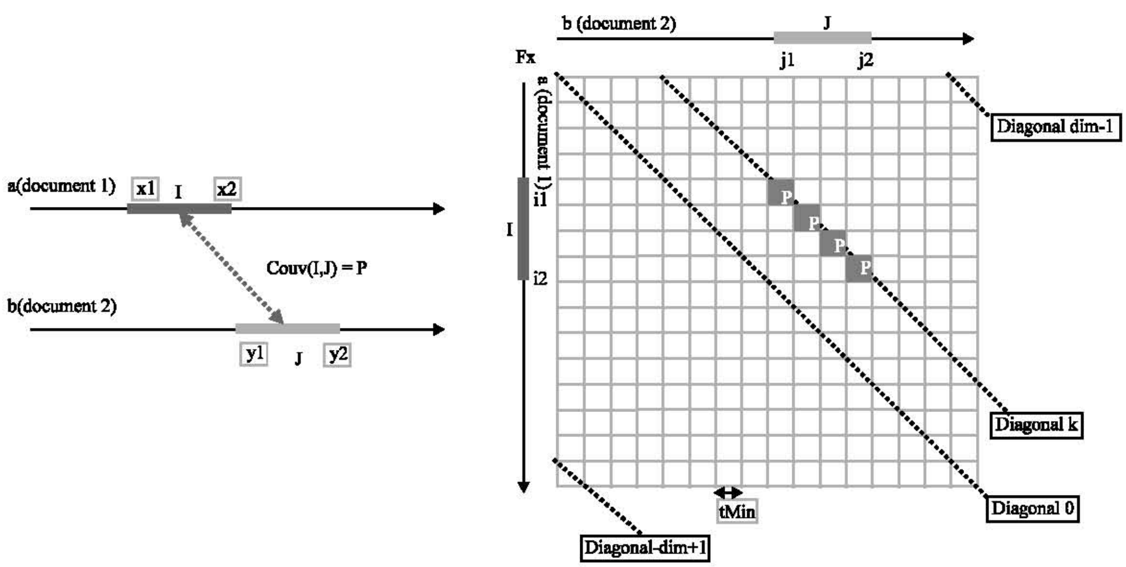

Fig. 1: Similarity matrix for the series a and b which represent the feature Fx of each of the documents (1 and 2) (Haidar *et al*., 2005)

Figure 1, we present a detailed example; We take two video documents: a(document1) and b(document2). For each document we extract a feature Fx, say the dominant color. This means that for each image of the video documents we calculated the dominant color, this color is represented by a value. The series extracted for a and b are time series. Figure 1, represents the comparison of Fx_a and Fx_b, following our matching method. In the figure we show two subsequences I (from $Fx_{a)}$ and J (from Fx_b). A subsequence of length 1 measures tMin images (tMin/25 sec). The matching for I and J measures a coverage P, thus, the shaded elements of the matrix receive the value P. They belong to the diagonal k, the diagonal zero being the main one. The fact that the sequence J is offset from the sequence I: k × tMin time units, k = (j1 - i1) means that if we superimpose a and b, the J segment is shifted (k×tMin) units time compared to the segment I.

So, in general for each pair of vectors (time series) representing an audiovisual feature, results of the comparison are recorded in a matrix, then a matrix merger is performed to extract the points of similarity common to all or most matrices in order to identify elements common to two videos.

Similarity matrix examples: The similarity matrix is interesting in the field of digital video as it provides a means of content analysis independent of the type of documents. Before analyzing the similarity matrix, we present some descriptive examples that give an overview on these matrices:

- Example 1: Auto-comparison: Comparing one commercial break with itself (Fig. 2).
- Example 2: Comparison: Comparing two different commercial breaks (Fig. 3)

Both matrices contain information comparing each pair of documents. Figure 2 is auto-comparing a video of length 1400×tMin, where tMin = 10, which is 2 h and 20 min. Also, Fig. 3 compares two different video of the same duration. One can easly notice the presence of dark points (hight similarity coverage) in the first matrix, whereas in the second one they are fewer.

To study the similarity of two particular scenes, the corresponding block to the scene could be simply extract and analyzed. This block is the matrix of similarity of the compared scenes.

We note that there are blocks or parts of warm colors (dark) and others of cool colors (light). In fact the intensity of the colors of the blocks is related to the value of the similarity measure between the compared segments. Dark colors indicate a high degree of similarity and light regions indicate a low similarity.

Figure 2 we note that there is a strong similarity on the diagonal, this since a scene is very similar to itself. Figure 3 we note that the similarities between the regions are dispersed throughout the matrix. In fact, the high values may well exist between different video segments if there is a similarity between these segments. Finally, we can still see areas of dissimilarity between different segments; these areas are of light colors.

In general a phase of post-treatment type smoothing and/or filter is required. The corresponding parameters are determined based on an analysis of the matrix.

AUTOMATIC MACRO-SEGMENTATION BASED ON REGIONAL MINIMUMS AND MAXIMUMS ANALYSIS

Regional minimum and maximum method: The macro-segmentation of the proposed method is obtained based the following 3-steps algorithm:

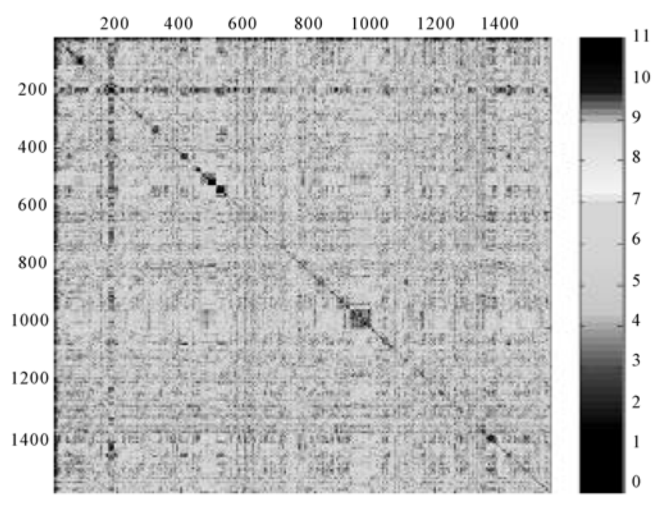

Fig. 2: Auto-similarity matrix

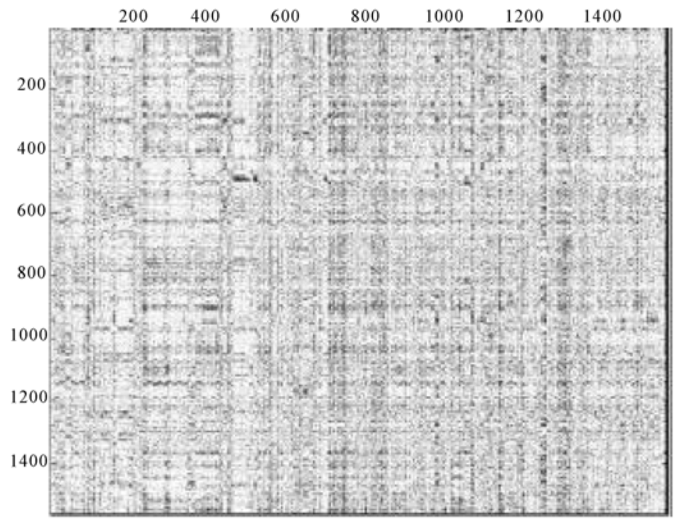

Fig. 3: Comparison matrix of two different video documents

The filtering step: The matrix is filtered twice using a Gaussian filter, each time with a different parameter. We apply a filter with a high parameter value followed by a second filter with a parameter of small value. The filter of low value will allow determining the regional minimum and that of high value will help determine the regional maximum while eliminating parasites values. It is necessary to make transformations for each value in the matrix, taking into account its neighborhood; this is done to fight against the noise effects.

The projection step: After filtering, the two matrices obtained will be projected once on the x-axis and another time on the y-axis, that is to say, after calculating the sum of the columns and the sum of lines.

The search step: After the projection of lines and columns of the matrix, the minimum and maximum regional projections are calculated. These minimums and maximums will play an important role in the separation regions. Considering the projections on the x-axis, for example, we look for the holes. Each local minimum, between two local maximums, is a hole. This hole means that the similarity between the two videos, according to the compared video characteristics, is minimal and consequently there is a change of regions. In practice this minimum was kept only if a threshold distance with surrounding maxima is reached. This threshold is chosen manually. Finally, we mark the limit and we draw a line between the two regions. For a self-comparison matrix, we can work only on either lines or columns since the matrix is symmetric.

EXPERIMENTATION

In this study, we present some of the macro-segmentation comparison matrices. We try to macro-segment commercial breaks recorded over two days of a French TV broadcasting, on an average of six break intervals per day. Summaries of these ranges are presented as thumbnails. Each thumbnail represents a commercial. The 62 commercials are then identified by numbers. A correlation graph is drawn manually to indicate the films shared between each pair of tracks. All this manual indexation and the commercial breaks database thumbnails are describe here.

Settings: We have macro segmented comparison matrices for each pair of breaks by taking the filter parameters of small values. We set the first sigma value, for the Gaussian filter, to 1.5 and the sec sigma value to 5.5 and the size of the filter window was set to 2, since most advertising is time-varying between 3 and 4 sec and since the value of the unit tMin is 16 (almost 3/4 of a sec). We finally set the threshold to 5. Recall that a second of video recording corresponds to 25 frames which is 25/tMin values (similarity measures) in the matrix.

RESULTS

We show two macro segmented matrices-each one is shown twice, before and after segmentation; one self-comparison matrix (Fig. 4) and one hybrid comparison matrix (Fig. 5). The white horizontal and vertical lines are places of separation between regions. These lines are

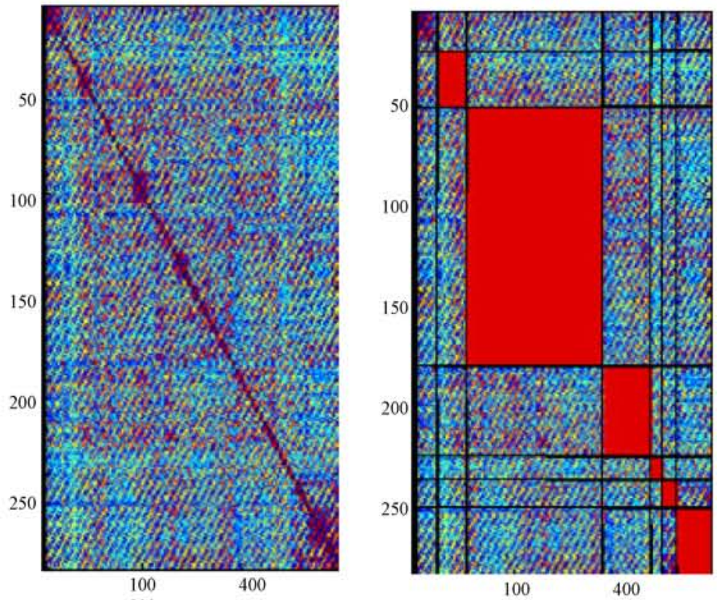

Fig. 4: Self-comparison matrix of Tuesday 3 (M3)

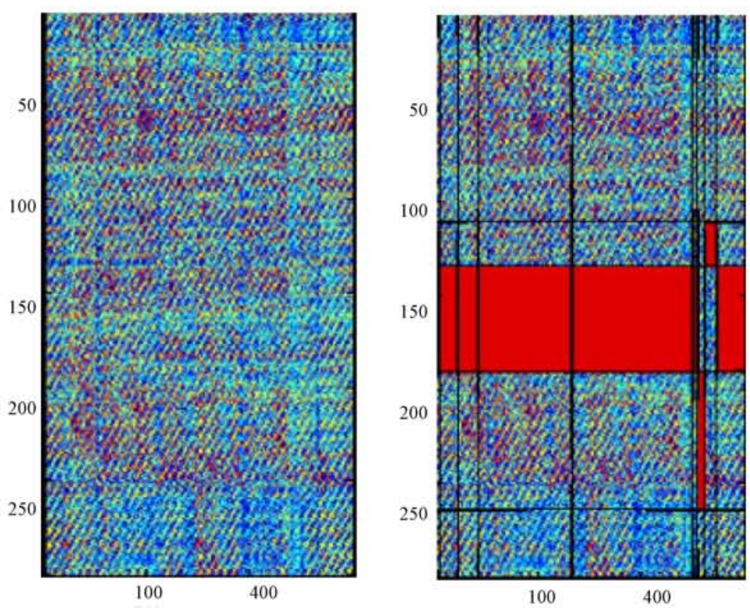

Fig. 5: Comparison matrix of Tuesday 3 and Monday 4 (M3 and L4)

deduced using our regional minimum and maximum method. Regions of bright red color are the places of the most similar. The right figure represents the main comparison matrix and the matrix on the left represents the macro segmented matrix.

In the below table we present in italic the results of the comparison matrix macrosegmentation between the commercials. The bold values are shared between the manual indexing and the macrosegmentation.

	L4M3	
M3M3	on x-axis	on y-axis
00:00:03	**00:01:11**	*00:00:02*
00:00:13	00:01:26	*00:00:14*
00:00:33	*00:02:01*	*00:00:27*
00:00:41	*00:02:27*	**00:01:25**
00:01:26		**00:02:27**
00:02:01		
00:02:12		
00:02:19		
00:02:32		
00:02:37		

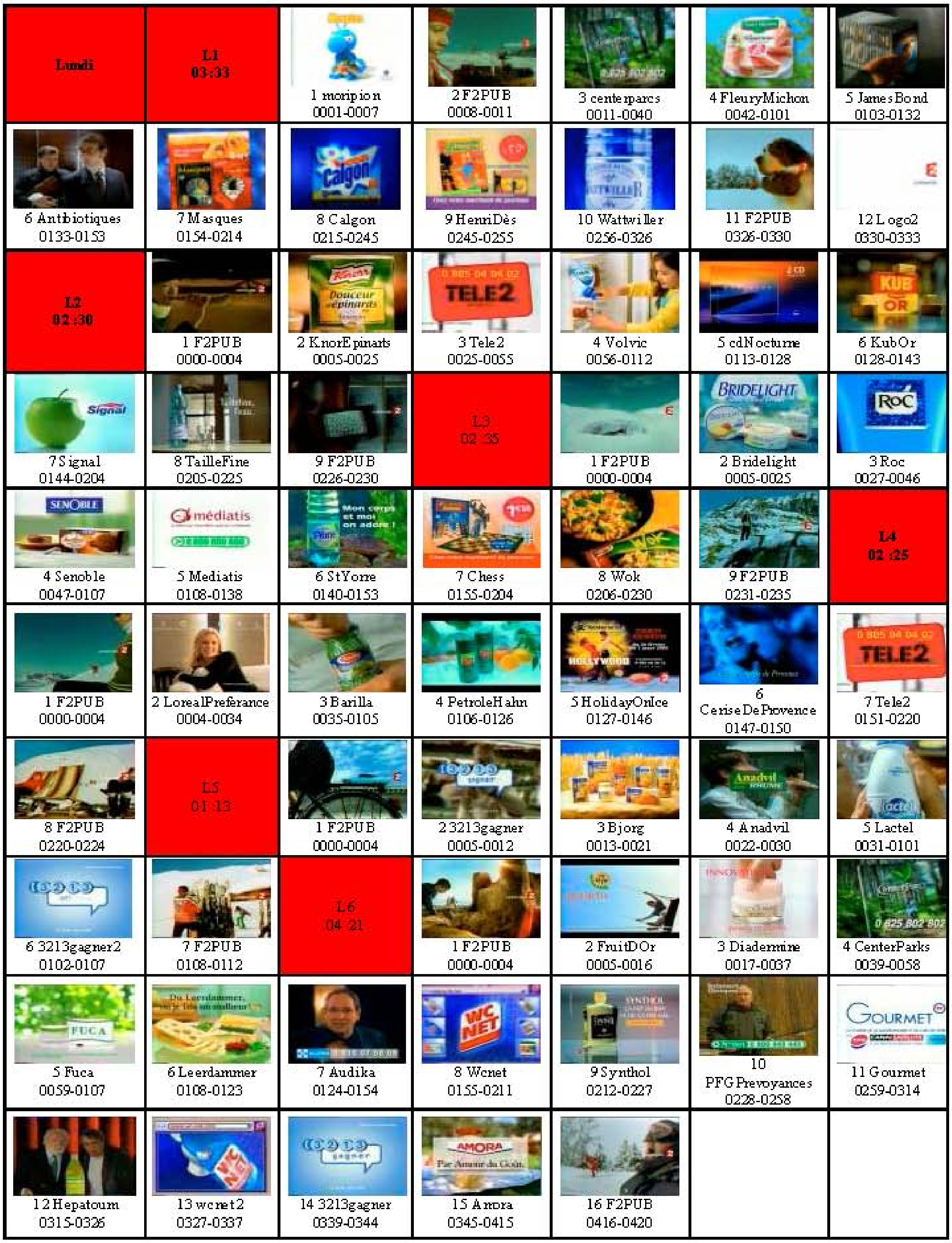

Fig. 6: The thumbnails of all the commercial breaks on Monday

Fig. 7: The thumbnails of all the commercial breaks on Tuesday

1	F2 PUB	22	Diadermine	43	Masques
2	3213 gagner	23	Dr. Pierre Richuad	44	Mediatis
3	AGagner	24	Eau Ecalrate	45	Moripion
4	Amora	25	Fleury Michon	46	Petrole Hahn
5	Anadvil	26	Florette	47	PFG Prevoyances
6	Antibiotiques	27	Fruit D'Or	48	ProPlan
7	Assurance Maladie	28	Fuca	49	Roc
8	Audika	29	Garnier Belle Color	50	Rothelec
9	BANature	30	Gourmet	51	Senoble
10	Barilla	31	Henri Des	52	Signal
11	Bjorg	32	Hepatoum	53	StYorre
12	Boursin	33	Holiday on ice	54	Synthol
13	Bridelight	34	James Bond	55	Taft
14	Brossard	35	Knor Epinards	56	Taille Fine
15	Calgon	36	KubOr	57	Tele 2
16	cdNocturne	37	Lactel	58	Volvic
17	Center Parks	38	Leerdammer	59	Wattwiller
18	Cerise de Privince	39	Libra	60	WCNet
19	Chess	40	Logo2	61	Weight watchers
20	Codotussyl	41	L'Oreal Preterence	62	Wok
21	Darty	42	L'Oreal Revitalift		

Fig. 8: Each film is given a unique identifier

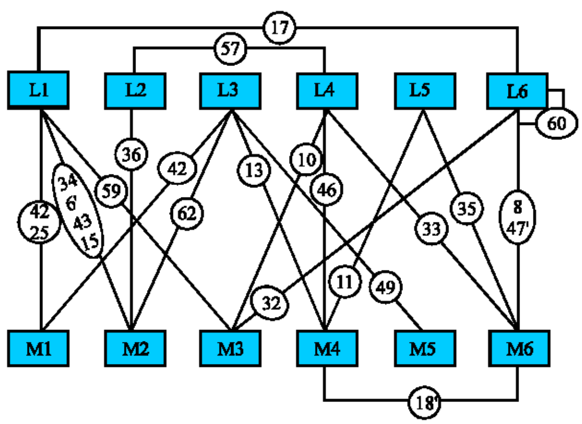

Fig. 9: Correlation graph between the commercial breaks. (L = Monday/M = Tuesday). The branches are labeled with the common films between nodes (break)

Here, we detail the commercial breaks database used in our experimentation. Figure 6 and 7 represent the thumbnails of all the commercial breaks. Each thumbnail is subnoted with the time interval during which it was broadcasted. Figure 8, we refer to each film by an identifier. And finally, in Fig. 9, we draw the graph of correspondence between the 12 breaks. The branches are labeled with the common films between nodes (break).

DISCUSSION

Before evaluating the method we must note that prior segmentation of documents depends on the tMin parameter that does not reflect the limitations of schemes, thus a difference of this amount between the manual segmentation and the method was tolerated. We define the following evaluation parameters:

- **Recall rate:** The number of relevant macro-segments selected divided by the number of relevant in macro-segments (s) document (s)
- **Accuracy or precision rate:** The number of relevant macro-segments selected divided by the number of selected macro-segments
- **Noise levels:** The difference between 1 and the accuracy rate
- **Rate of silence:** The difference between 1 and the recall rate

We can calculate the recall and precision rate.

	M3M3	L4M3	Average rates for ALL the commercials
Recall rate	4/10 = 0.4	1/1 = 1	0.65
Precision rate	4/10 = 0.4	1/4 = 0.25	0.47

Present method yielded macro-segmentation compared with the limits of manually indexed references to three types of results: false detections, missing detections and correct detections.

The incorrect detections have the effect of decreasing the accuracy rate since they increase the number of incorrect macro-segments. Missing detections therefore also reduce the precision and recall rates because they increase the number of incorrect macro-segments and reduce the number of correct results.

We interpret the presence of false detections and missing detections by the following causes; on one hand, the values of the comparison matrix depend on the parameter tMin, which features visual basis for comparison and fusion of these characteristics. A change in behavior or a subset of visual features within a single macro-segment is one of the causes of a false detection. On the other hand, the parameters of filtering and threshold selected, play a key role influencing the results. These filter parameters and threshold can be improved and even the type of filter chosen, may be changed.

These results have wide potential and demonstrate the effectiveness of our method.

CONCLUSIONS

We proposed a similarity matrix macro-segmentation method based on the search for local maxima and minima indicating regional changes. We presented an application example based on commercial breaks and finally we evaluated our method by precision and recall rates.

Finally we propose a set of tracks inspired by work that could be examined in detail. Regarding the filter, we can improve results by changing filter settings or by finding another more appropriate filter that can remove the noise effects without affecting the similarity values. We may also consider the threshold values used to decide whether there is a local minimum between two maximums.

REFERENCES

Adams, B., C. Dorai and S. Venkatesh, 2000. Novel approach to determining tempo and dramatic story sections in motion pictures. Proceedings of the 2000 International Conference on Image Processing, Sept. 10-13, Vancouver, BC, Canada, pp: 283-286.

Aigrain, P., P. Joly and V. Longueville, 1997. Medium Knowledge Based Macro-Segmentation of Video into Sequences. In: Intelligent Multimedia Information Retrieval, Maybury, M.T. (Ed.). MIT Press, Cambridge, MA, USA., ISBN: 0-262-63179-2, pp: 159-173.

Ariki, Y. and Y. Saito, 1996. Extraction of TV news articles based on scene cut detection using dct clustering. Proceedings of the International Conference on Image Processing, Sept. 16-19, Lausanne, pp: 847-850.

Carrive, J., F. Pachet and R. Ronfard, 2000. Clavis: A temporal reasoning system for classification of audiovisual sequences. Proceedings of the Intenational Conference on Content-Based Multimedia Access (RIAO), April 12-14, Paris, France, pp: 1400-1415.

Collins, J.F. and A.F. Coulson, 1990. Significance of protein sequence similarities. Methods Enzymol., 183: 474-487.

Cooper, M. and J. Foote, 2001. Scene boundary detection via video self-similarity analysis. Proceedings of the 2001 IEEE International Conference on Image Processing, Oct. 7-10, Thessaloniki, Greece, pp: 378-381.

Daget, J. and J.R. Durand, 1968. Étude du peuplement de poissons d'un milieu saumâtre tropical poïkilothermie: La baie de cocody en côte d'ivoire, Calz. ORSTOM Ser. Hydrobiol., 2: 91-111.

Feng, Y., R. Ren and J.M. Jose, 2008. Rule-based scene boundary detection for semantic video segmentation. Proceedings of the 5th International Conference on Visual Information Engineering (VIE), July 29-Aug. 01, Xian China, pp: 667-672.

Galmar, E., T. Athanasiadis, B. Huet and Y. Avrithis, 2008. Spatiotemporal semantic video segmentation. Proceedings of the 10th Workshop on Multimedia Signal Processing, Oct. 8-10, Cairns, Qld, pp: 574-579.

Gunsel, B., A.M. Ferman and A.M. Tekalp, 1998a. Temporal video segmentation using unsupervised clustering and semantic object tracking. J. Electronic Imaging, 7: 592-604.

Gunsel, B., M. Tekalp and P. Van Beek, 1998b. Content-based access to video objects: Temporal segmentation, visual summarization and feature extraction. Signal Process., 66: 261-280.

Haidar, S., P. Joly and B. Chebaro, 2004. Audiovisual production invariant searching. CORIA, 4: 333-346.

Haidar, S., P. Joly and B. Chebaro, 2005. Style Similarity Measure for Video Documents Comparison. In: Image and Video Retrival, Leow, W.K. (Ed.). Springer Verlag, Berlin, ISBN: 978-3-540-27858-0, pp: 307-317.

Haidar, S., P. Joly and B. Chebaro, 2006. Mining for video production invariants to measure style similarity. Int. J. Intel. Syst., 21: 747-763.

Joly., P. and P. Aigrain, 1996. Consultation et analyse des documents en image animée numérique. Ph.D. Thesis, Travaux Universitaires - Thèse nouveau doctorat, pp: 342. http://cat.inist.fr/?aModele=affiche N&cpsidt=175608.

Lienhart, R., S. Pfeier and W. Eelsberg, 1999. Scene determination based on video and audio features. Proceedings of the International Conference on Multimedia Computing and Systems, June 7-11, Florence, Italy, pp: 685-690.

Merlino, A., D. Morey and M. Maybury, 1997. Broadcast news navigation using story segmentation. Proceedings of the 5th ACM International Conference on Multimedia, Nov. 9-13, Seattle, Washington, United States, pp: 381-391.

Peyrard, N. and P. Bouthemy, 2002. Content-based video segmentation using statistical motion models. Proceedings of the International Conference on British Machine Vision, (BMVC'02), Cardiff, France, pp: 527-536.

Rui, Y., T. Huang and S. Mehrotra, 1999. Constructing table-of-content for videos. Multimedia Syst., 7: 359-368.

Saraceno, C. and R. Leonardi, 1997. Audio as a support to scene change detection and characterization of video sequences. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, April 21-24, Munich, Germany, pp: 2597-2600.

Sickle, J., 1997. Using mean similarity dendograms to evaluate classifications. J. Agric. Biol. Environ. Statistics, 2: 370-388.

Sivaselvan, B. and N.P. Gopalan, 2006. An efficient frequent temporal pattern mining algorithm. Inform. Technol. J., 5: 1043-1047.

Yahya, A.A. and A.M. Abdalla, 2003. Implementation of matrix folding in multimedia transmission. Inform. Technol. J., 2: 154-158.

Yeung, M., B.L. Yeo and B. Liu, 1998. Segmentation of video by clustering and graph analysis. Comput. Vision Image Understanding, 71: 94-109.

Zhang, H., Y. Gong, S. Smoliar and S. Tan, 1994. Automatic parsing of news video. Proceedongs of the International Conference on Multimedia Computing and Systems, May 14-19, Boston, Massachusetts, pp: 45-54.