High-Quality Real-time Temporal Segmentation Tool for Video Editing Software

Carlos Cuevas and Narciso García

***Abstract*— The increasing use of video editing software has resulted in a necessity for faster and more efficient editing tools. Here, we propose a lightweight high-quality video indexing tool that is suitable for video editing software.**

I. INTRODUCTION

In our days, video editing software is commonly used by professional and non-professional users [1]. This software includes tools for index and retrieve relevant material that, in a first step, perform a temporal segmentation in shots [2].

To provide high-quality results, these tools apply complex strategies that are not fast enough and, additionally, depend on a significant amount of thresholds that should be adequately fixed by the users [3].

Here, we propose a novel real-time high-quality shot detection strategy, which is suitable for video editing software requiring both, low computational cost and high Recall and Precision percentages. While abrupt transitions are detected though a very fast pixel-based analysis, gradual transitions are obtained from an efficient edge-based analysis. Both analyses are reinforced with a motion analysis that is carried out exclusively over a reduced amount of candidate transitions, then, maintaining the computational requirements.

II. ABRUPT TRANSITION DETECTION

Frames belonging to a single shot are more similar than frames belonging to different shots. Then, abrupt transitions can be efficiently detected though the computation of the differences between pixel intensity values of consecutive images. Then, we propose a novel and powerful metric that allows detecting most of the abrupt transitions and avoids false detections resulting from illumination changes:

$$M\left(I^{n}\,|\,I^{n-1}\right)=\frac{1}{HW}\sum_{h,w}\rho_{h,w}^{n}$$

where I^n and I^{n-1} are the compared consecutive images, (H,W) are the image dimensions, (h,w) are pixel coordinates, and:

$$\rho_{h,w}^{k}=\begin{cases}1, \text{if } \operatorname{sign}\left(I_{h,w}^{n}-\mu^{n}\right)=\operatorname{sign}\left(I_{h,w}^{n-1}-\mu^{n-1}\right)\wedge\left|I_{h,w}^{n}-\mu^{n}\right|>T_a\\ -1, \text{if } \operatorname{sign}\left(I_{h,w}^{n}-\mu^{n}\right)\neq\operatorname{sign}\left(I_{h,w}^{n-1}-\mu^{n-1}\right)\wedge\left|I_{h,w}^{n}-\mu^{n}\right|>T_a\\ 0, \text{otherwise}\end{cases}$$

where μ_n is the mean intensity of I^n and T_a is a noise threshold.

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under project TEC2010-20412 (Enhanced 3DTV).

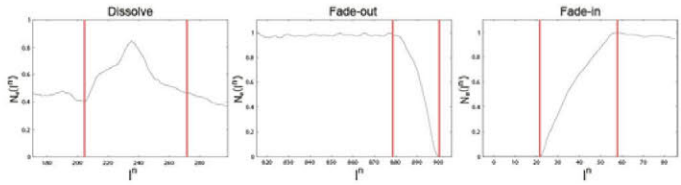

Fig. 1. Evolution of the edge-point amount along different kind of gradual transitions.

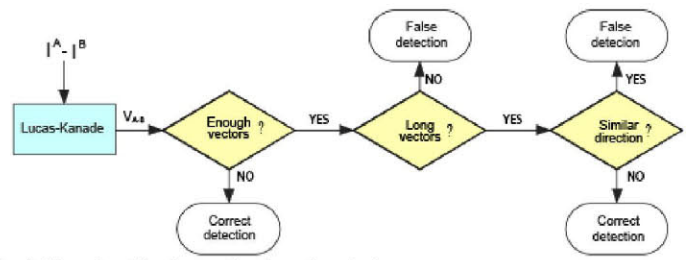

Fig. 2. Flowchart for the motion-based analysis.

III. GRADUAL TRANSITION DETECTION

In a gradual transition, the first shot edges gradually disappear while the second shot edges gradually appear. Then, the gradual transitions can be detected though the analysis of the evolution of the amount of edges, $N_e(I^n)$ along the sequences. Fig. 1 depicts some examples of the evolution of the amount of edges along different kind of gradual transitions.

IV. MOTION-BASED PRUNNING

Previously described strategies are able to detect most abrupt and gradual transitions in real-time. However, they do not avoid some false detections resulting from undesirable situations such as fast camera displacements, zooms, etc. To detect and separate these false detections from the correct ones, we propose an efficient and innovative motion analysis applied over the previously detected transitions, resulting from the pixel-based and edge-based algorithms.

Usually, motion-based analyses provide high quality detections (high Recall and Precision percentages), but they are computationally inefficient [4]. Nevertheless, the proposed motion analysis is carried out exclusively over a reduced amount of candidate transitions and, consequently, the computational requirements of the proposed strategy are maintained.

The flowchart of the proposed motion-based strategy is detailed in Fig. 2. Applying a Lucas-Kanade pyramidal algorithm [5] over each pair of images (I^A-I^B) delimiting each candidate transition, a set of motion vectors, $\{v_1 \ldots v_N\}$, linking a set of singular points in I^A, $\{s_1 \ldots s_M\}$, with points in I^B is

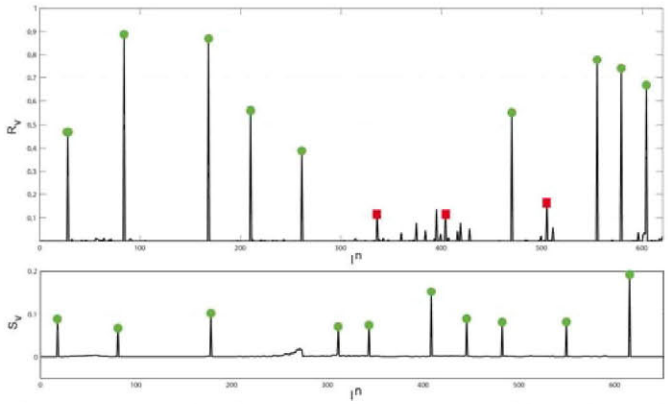

Fig. 3. (a) Amount of motion vectors along a sequence with 615 frames. (b) Mean motion vector length along a sequence with 650 frames. Green circles mark real transitions and red squares mark false detections.

obtained. Firstly, the ratio $R_v=1-v_N/s_M$ is computed. As it is shown in the top graphic depicted in Fig. 3, images belonging to a same shot result in much lower R_v values that images from different shots. Then, this ratio allows to discriminate between correct detections and some false detections. In a second step, to discard false detections resulting from large moving objects and camera motion, the mean length of the vectors, S_v, is analyzed. Between images from a same shot, this length is significantly lower than between images from different shots (bottom graphic in Fig. 3). Finally, to identify false detections resulting from fast camera changes (traveling, pans, tilts or zooms), we analyze the typical deviation of the set of vectors:

$$\sigma_v = \left(\sum_{i=1}^{v_n} L_i \right)^{-1} \sum_{i=1}^{v_N} L_i (D_i - \mu_v)^2$$

where L_i is the length of the *i-th* vector, D_i is its direction, and μ_v is the mean direction of the set of motion vectors. Abovementioned camera changes produce false detections where most motion vectors have similar orientations. Consequently, these false detections result in significantly lower σ_v values than correct detections and, thus, they can be easily discarded.

V. RESULTS

Table I and II present the Recall and Precision percentages obtained with the proposed strategy over approximately three hours of video sequences with more than *950* transitions. These results show that, through the pixel-based and edge-based analyses, most abrupt and gradual transitions are detected (high Recall values). Moreover, with the proposed motion analysis most false detections are discarded (high Precision values).

On the other hand, to show the low computational cost of the proposed strategy, Table III shows some obtained frame rates, before and after the motion-based analysis, for sequences with different spatial resolutions.

TABLE I
RECALL AND PRECISION PERCENTAGES OBTAINED WITH THE PROPOSED ABRUPT TRANSITION DETECTION STRATEGY

Sequence	Abrupt transitions	Without motion analysis		With motion analysis	
		Recall	Precision	Recall	Precision
Cartoons	279	100	84.80	98.21	94.48
Musicals	206	99.03	92.73	97.09	98.04
Reports	82	100	92.13	92.68	100
Other	284	100	91.61	99.65	100
Total	851	99.76	89.56	97.88	97.66

TABLE II
RECALL AND PRECISION PERCENTAGES OBTAINED WITH THE PROPOSED GRADUAL TRANSITION DETECTION STRATEGY

Sequence	Gradual transitions	Without motion analysis		With motion analysis	
		Recall	Precision	Recall	Precision
Cartoons	7	85.71	35.29	85.71	66.67
Musicals	8	100	47.06	100	72.73
Reports	77	93.51	79.12	92.21	95.95
Other	1	100	7.14	100	33.33
Total	93	93.55	62.59	92.47	88.66

TABLE III
OBTAINED FRAME RATE (FRAMES PER SECOND)

Sequence	Dimensions (HxW)	Before the motion analysis	After the motion analysis
Cartoons	182x320	61.59	59.76
Musicals	270x480	56.31	55
Reports	360x480	45.82	45.59
Other	360x640	32.08	30.43

VI. CONCLUSIONS

Here, a novel lightweight high-quality shot detection strategy has been described. Applying pixel-based and edge-based algorithms, candidate transitions are obtained. Finally, with an efficient motion analysis over these transitions, false detections are separated from the correct ones.

Obtained results have shown that the proposed strategy works very fast and provides high Recall and Precision percentages in a large variety of sequences. Then, it can be perfectly used in video editing software tools where speed and quality are required by users.

REFERENCES

[1] W. T. Peng, W. T. Chu, C. H. Chang, C. N. Chou, W. J. Huang, W. Y. Chang, and Y. P. Hung, "Editing by Viewing: Automatic Home Video Summarization by Viewing Behavior Analysis", IEEE Trans. Multimedia, vol. 13, no. 3, pp. 539-550, 2011.

[2] A. Smeaton, P. Over, and A. Doherty, "Video shot boundary detection: Seven years of TRECVid activity", Computer Vision and Image Understanding, vol. 114, no. 4, pp. 411-418, 2010.

[3] D. Brezeale, and D. Cook, "Automatic video classification: A survey of the literature", IEEE Trans. Systems Man and Cybernetics-C, vol. 38, no. 3, pp 416-430-2008.

[4] Y. Gao, W. B. Wang, and J. H. Yong, "A video summarization tool using two-level redundancy detection for personal video recorders", IEEE Trans. Consumer Electronics, vol. 54, no. 2, pp. 521-526, 2008.

[5] J. Bouguet, "Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm", Intel Corporation, Microprocessor Research Labs, OpenCV documents, 1999.