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:

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

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

\[ \rho_{h,w} = \begin{cases} 1, & \text{if } \text{sign}(I^n_{h,w} - \mu^n) = \text{sign}(I^{n-1}_{h,w} - \mu^{n-1}) \land |I^n_{h,w} - \mu^n| > T_a \\ -1, & \text{if } \text{sign}(I^n_{h,w} - \mu^n) \neq \text{sign}(I^{n-1}_{h,w} - \mu^{n-1}) \lor |I^n_{h,w} - \mu^n| > T_a \\ 0, & \text{otherwise} \end{cases} \]

where \( \mu^n \) is the mean intensity of \( I^n \) and \( T_a \) is a noise threshold.

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, \( N_e(t) \) 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 PRUNING
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 \((I^1, I^2)\) delimiting each candidate transition, a set of motion vectors, \( \{v_1, \ldots, v_M\} \), linking a set of singular points in \( I^1, \{s_1, \ldots, s_M\} \), with points in \( I^2 \) is
obtained. Firstly, the ratio $R_v = 1 - v_S / S_M$ 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, $S_M$, 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:

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

where $L_i$ is the length of the $i$-th vector, $D_i$ is its direction, and $\mu_v$ 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 $\sigma_v$ 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.

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


