IMPROVED CUT DETECTION FOR THE SEGMENTATION OF ANIMATION MOVIES

B. Ionescu, V. Buzuloiu

University Politehnica Bucharest
LAPI, 061071 Romania
bionescu,buzuloiu@alpha.imag.pub.ro

P. Lambert, D. Coquin

Universite de Savoie, LISTIC
B.P. 806, 74016 Annecy, France
patrick.lambert,didier.coquin@univ-savoie.fr

ABSTRACT

In this paper an improved cut detection algorithm, adapted to the segmentation of animation movies, is proposed. As color is a major feature of animation movies (each movie has its own particular color distribution) the proposed algorithm applies second order derivatives on Euclidean distances between color histograms of frames quadrants in order to improve the cut detection. For the frame classification, an automatic threshold estimation is proposed. Also, in order to reduce false detections, we propose an algorithm to detect an effect specific to animation movies, named ”short color change” (i.e. thunders, lightening). The resulting method achieved better results compared to the classical histogram–based and motion–discontinuity based approaches, as shown by tests conducted on several animation movies.

1. INTRODUCTION

As multimedia content became increasingly accessible owing to the development of new video coding techniques and of devices with bigger storage capabilities, the interest on video databases indexation has drastically increased. Thanks to the “The International Animated Film Festival” [1], which takes place every year at Annecy, France since 1960, a very large database of animation movies is available. Managing thousands of videos is a tedious task; therefore an automatic content analysis would be more than welcome.

Detecting the video shots boundaries, that is, recovering the elementary video units, provides the ground for nearly all existing video abstraction and high–level video segmentation algorithms [2].

In this paper, an improved, cut detection technique is proposed to cope with the issues raised by the peculiarity of animation movies. Animation movies are different from natural ones in that: the events do not follow a natural way, the camera motion is very complex, the characters usually are not human and could have any shape, a lot of visual/color effects are used, every animation movie has its own particular color distribution, artistic concepts are used (i.e. painting concepts), various animation techniques are used (3D, cartoons, animated objects, etc.).

Existing cut detection algorithms differ in the features used to measure the visual discontinuity. Cut detection techniques can be classified as: intensity/color based, edge/contour based and motion based [2]. A comparison of all of the three approaches is presented in [2][3]. In the class of intensity/color–based methods, the histogram–based methods achieve better results comparing to the other methods [4]. Various approaches to histogram–based cut detection were proposed using color based histogram metrics or histogram differences in several color spaces (for a literature survey see [5]).

The proposed cut detection algorithm exploits the color feature, as being the major feature of animation movies. The frames are color reduced and divided into four quadrants in order to reduce the influence of entering objects (as the predominant motion in animation movies is the object motion [5]). A study on the influence of objects size on the global color histogram is performed. By using second order derivatives computed on a temporal mean of the Euclidean histogram distances, the influence of the repetitive camera motion is reduced and cuts better emphasized. False detections are reduced by detecting an effect specific to animation movies, which is called “short color change” or SCC (i.e. thunders, explosions) which is usually (wrongly) detected as a cut. The proposed algorithm uses a modified flash–detector. For frame classification, an automatic threshold estimation is proposed using statistical measures computed on the obtained distance vectors. The results obtained by the proposed method are better compared to the classical histogram–based and motion discontinuity–based cut detection approaches.

The remainder of the article is organized as following: Section 2 describes the proposed cut detection algorithm, in Section 3 we present some experimental results and Section 4 contains final considerations and future improvements.

2. CUT DETECTION

2.1. Video subsampling and color reduction

In order to reduce computational complexity, the video sequence is first temporally subsampled: only one frame in n is retained for further processing and second spatialy subsampled: only one pixel for each block of 4×4 pixels is retained (see Section 3). In order to compute color histograms, the number of colors are reduced using the Floyd–Steinberg error diffusion algorithm run in the XYZ color space, by using a standard color palette: the 216 webmaster color palette, as described in [6]. For a study on the influence of the color reduction on the cut detection see [7].

2.2. Color histogram computation

Animation movies have some peculiarities that must be specifically addressed. One of the most important issues when dealing with the color histogram is the movement of objects in the scene, as large–sized moving objects may produce noticeable differences in the histograms of successive frames. The predominant type of motion in animation movies is the object motion [5]. A study conducted on the influence of the object's size on the global color histogram has shown that only the objects of the size of an image quadrant or higher will significantly change the global color histogram (see Section 3). In order to reduce the method's sensitivity to emerging/vanishing objects, frames are divided into quadrants.

For each retained frame f_t^c (where ^{c} denotes the subsampled version and t the time index) four color histograms $H_t^j = H^j(f_t^c, i)$ (with i indexing colors) are computed, each one corresponding to an image quadrant (indexed by j). Then four Euclidean distances, $D_j(t)$, between each color histogram in the frame f_t^c and the corresponding one in the next temporal frame f_{t+1}^c are computed as:

$$D_j(t) = \left(\sum_{i=1}^{N_c} \left[H^j(f_{t+1}^c, i) - H^j(f_t^c, i) \right]^2 \right)^{1/2} \qquad (1)$$

where $N_c = 216$ represents the number of colors of the chosen palette. An average histogram distance $D_{mean}()$, between consecutive frames is computed as the arithmetic mean of the four $D_j(t)$ differences to stand as the basis for the cut detection procedure we shall explain in the sequel.

The advantage of using a single frame difference measure instead of four is twofold. It firstly leads to the simplification of the cut/non–cut decision. Secondly, thorough tests have shown that, provided that an accurate threshold selection procedure is devised (see Section 2.4), the method based on $D_{mean}(t)$ measure leads to improved results.

2.3. Second order derivative

A cut means a strong color dissimilarity between two consecutive frames (i.e., a high value of the $D_{mean}(t)$ value) that continues with a strong color similarity between frames (i.e., a low value of $D_{mean}(t)$). A simple temporal gradient computed on the $D_{mean}(t)$ sequence does not take into account more than a pair of neighbors, thus higher–order derivatives should be used for accurate cut detection.

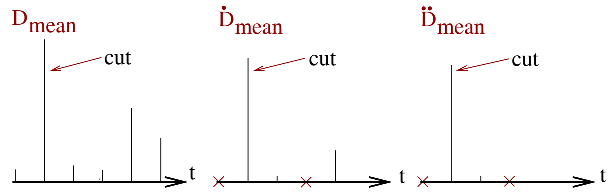


Fig. 1. Cuts emphasizing using second order derivatives. The oX axis corresponds to time, the oY axis corresponds to the D_{mean}, $\dot{D}_{mean}$ and respectively $\ddot{D}_{mean}$ sequences, negative values are set to 0 (marked with the red $\times$).

Tests performed using p–order derivatives have shown that the higher the p value used, the lower the obtained cut detection rate (the distance values decline as p increases). The use the second order derivative has led to the best compromise between detection rate and false positive incidence. Moreover, the second order derivative is the best match for our problem, which consists in detecting the "low–high–low" pattern in the D_{mean} sequence (see Figure 1). Thus, the second order derivative of $D_{mean}()$ is computed as:

$$\ddot{D}_{mean}(t) = \frac{\partial^2 D_{mean}(t)}{\partial t^2} \qquad (2)$$

where t is the frame index. Negative values are set to 0 as they contain redundant information. Cuts are then sought as local maxima of the $\ddot{D}_{mean}$ sequence.

The advantage of using the second order derivative is illustrated in Figure 2 on a sample of an animation movie containing repetitive camera motion. The repetitive camera motion is the most frequent cause for false positives as it induces significant differences between histograms of consecutive frames. However, by using our second–derivative–based approach, the influence of repetitive camera motion is drastically reduced and real cuts are better emphasized.

2.4. Automatic threshold estimation

For the final frame classification, color histogram mean distances are compared to a certain threshold s. Using a global pre–fixed threshold s is practically impossible, as each animation movie has its own color distribution [8]. Several approaches to the estimation of the classification threshold are discussed in [2]. The most frequently used threshold determination is based on the average $\ddot{D}_{mean}$ value, $m_{\ddot{D}}$. As cuts occurrence is reduced, taking the threshold as $m_{\ddot{D}}$ would lead to a high false detection ratio (the threshold is too low). Our approach to threshold determination is different. The proposed threshold is determined in two steps by detecting all the significant local maxima of the $\ddot{D}_{mean}(t)$ sequence, and then the threshold s is set to the average value of the retained peaks. Experimental result have proven that this choice leads to a very good detection rate (see Section 3).

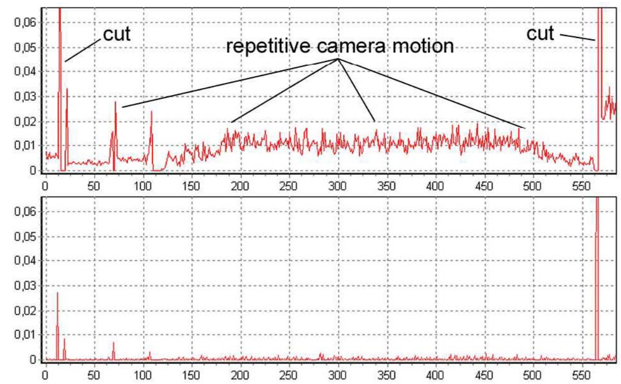


Fig. 2. Improvements to the cuts delimitation by using the second derivative: the $D_{mean}(t)$ sequence (top) and the $\ddot{D}_{mean}(t)$ sequence (bottom).

Thus, the cut detection is performed by thresholding the $\ddot{D}_{mean}$ sequence with the proposed threshold s. More precisely, a cut is detected whenever:

$$\ddot{D}_{mean}(t) > s \quad \text{AND} \quad \ddot{D}_{mean}(t-1) < s \qquad (3)$$

2.5. False detection reduction

Because the animation movies contain a lot of special visual effects, the occurrence of false cuts are very likely. In order to reduce the false detection rate, every detected cut is additionally checked in view of detecting a color effect specific to animation movies called "short color change" or SCC (see Figure 3), effect that is responsable for a part of the false positives. The proposed algorithm is inspired from the flashlight detection in natural movies [9].

Fig. 3. Examples of "short color changes" from the movie "Francois le Vaillant" [1].

As a SCC starts with an important change in color and ends with almost the same frame as the starting one. Distances between histograms $H(f_t^c)$ and $H(f_{t+1+l}^c)$ are successively computed, where f_t^c is the retained frame subsampled and color reduced at time index t, $l = 1, 2, 3, \ldots L_{\max}$ with $L_{\max}$ the maximal admitted length of the SCC effect, set to 10. If, for a given l, the distance between the two histograms is lower than s (with s being the threshold used in cut detection, defined in Section 2.4) the detected cut is classified as SCC, and gets, thus, discarded.

3. EXPERIMENTAL RESULTS

For the choice of the temporal subsampling step, set to $n = 2$, several tests were performed for different values of n ($n \in \{1, .., 10\}$). Also, an adaptive subsampling proposed in [10] was tested. The adaptive subsampling uses a "divide et impera" based algorithm: the sequence is first subsampled using a high step value and if a cut occurred the step is divided progressively in order to localize the cut precisely. However, the procedure proved too time consuming, especially for movies with short shots (i.e., animation movies) and it was abandoned. In what concerns the splitting of the frame into subframes in order to compute the local histograms, a study on the influence of different object sizes on the global color histogram proved that only objects of the size of an image quadrant or higher change significantly the global color histogram, leading to false detections. Thus the frames were divided into only four quadrants.

3.1. Detection results

The performance of the proposed methods are evaluated by using the precision/recall ratios, defined as:

$$\text{precision} = \frac{\text{GD}}{\text{GD} + \text{FD}}, \quad \text{recall} = \frac{\text{GD}}{\text{N}_{\text{total}}} \qquad (4)$$

where GD is the number of good detections, FD is the number of false detections and N_{total} is the total number of real transitions. The SCC detection was validated using 14 short–time animation movies from [1] with total time of 101min47s (containing 120 SCC) and achieved an overall precision and recall of **93%** and **88.3%** respectively.

The proposed cut detection algorithm was tested on two long animation movies with a total time of 158min and 3166 cuts (movie1: 85min and 1597 cuts and movie2: 73min and 1569 cuts). The results are compared with the classical histogram based method (referred as "*histogram method*") which uses directly the frames color histograms in order to detect the cuts and with a motion based method (referred as "*motion method*") which uses a motion discontinuity–based approach.

In order to estimate the detection errors, cuts have been manually labeled using a specially–developed software. The detection results are presented in Table 1 and Figure 4. The obtained false detections are mostly owing to the very fast camera motion, while misdetections to color similarities between the cut's frames. Some examples of the most frequent error situations are illustrated in Figure 5.

The *histogram method* leds to a lower recall, 88.63%, thus a lower recognition rate, while the *motion method* achieved a better recognition but with a higher false detection ratio, precision 89.39%.

Using the proposed method we have obtained both higher precision and higher recall ratios (above **92%**, see Figure 4). The recall obtained with the *histogram method* was improved

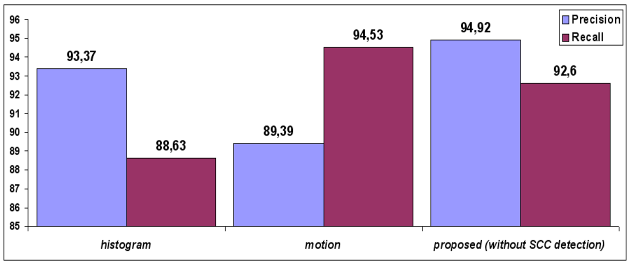


Fig. 4. The obtained cut detection precision and recall.

Method	GD	FD	precision	recall
histogram	2806	199	**93.37%**	**88.63%**
motion	2993	355	**89.39%**	**94.53%**
proposed	2931	157	**94.92%**	**92.6%**
proposed$^+$	2931	127	**95.97%**	**92.6%**

Table 1. The obtained cut detection errors (see equation 4, with $N_{total} = 3166$, $^+$ denotes the inclusion of the SCC detection).

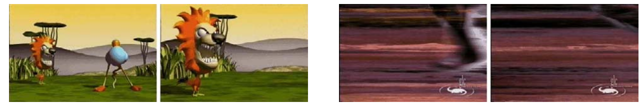
of **4%** (125 cuts) while the precision of the *motion method* was improved of **5.5%** (178 cuts). Also, using the proposed method in conjunction with the SCC detection procedure, the precision was up to **96%** (improved of 1%, see Table 1).

4. CONCLUSIONS

In this paper an improved cut detection technique based on second-order derivatives of Euclidean distances, adapted to the segmentation of animation movies is proposed. A number of improvements are used in order to manage the specificities of animation movies, which aim at reducing the influence of object/camera motion and of special visual effects. For the frame classification an automatic threshold estimation is proposed.

The proposed method achieved a very good recognition ratio compared with the classical histogram–based and motion–based approaches. The obtained recall and precision are above 92%, that is an improvement of around 3% (more than 100 cuts). Also, by using the proposed method in conjonction with the "short color change" detection, the precision rate was improved of 1%.

Future improvements consists on merging the detection results with the ones obtained with the *motion method*, all in order to reduce the misdetections. A possible strategy is based on the use of a confidence measure in order to describe the quality of the detection for both the two methods. Also a multimodal analysis is to be considered by analysing the movie's sound track (particular sound configurations are related to cuts).



(a) frames 2329,2328 (b) frames 4923,4924

Fig. 5. Cut misdetection/false detection examples: (a) color similarity (misdetection), "Gazoon" movie; (b) very fast motion (false detection), "The Buddy System" movie, see [1]).

5. REFERENCES

[1] Centre International du Cinema d'Animation, "http://www.annecy.org".

[2] R. Lienhart, "Reliable Transition Detection in Videos: A Survey and Practitioner's Guide" *International Journal of Image and Graphics*, Vol. 1(3), pp. 469-486, 2001.

[3] A. Dailianas, R. B. Allen, P. England, "Comparison of Automatic Video Segmentation Algorithms" *SPIE Integration Issues in Large Commercial Media Delivery Systems*, Vol. 2615, pp. 2-16, 1995.

[4] U. Gargi, R. Kasturi, S. H. Strayer, "Performance Characterization and Comparison of Video-Shot-Change Detection Methods" *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10(1), 2000.

[5] Cees G.M. Snoek, M. Worring, "Multimodal Video Indexing: A Review of the State-of-the-art", *Multimedia Tools and Applications*, in press, 2005.

[6] B. Ionescu, D. Coquin, P. Lambert, L. Dârlea, "Color-Based Semantic Characterization of Cartoons", *IEEE ISSCS - International Symposium on Signals, Circuits and Systems*, Iasi, Romania, Vol. 1, pp. 223-226, 2005.

[7] B. Ionescu, D. Coquin, P. Lambert, V. Buzuloiu, "The Influence of the Color Reduction on Cut Detection in Animation Movies", *Proceedings of* 20^{eme} *Colloque GRETSI sur le Traitement et l'Analyse du Signal et d'Image*, Louvain-la-Neuve, Belgique, september 2005.

[8] R. Lienhart, "Comparison of Automatic Shot Boundary Detection Algorithms" *SPIE Storage and Retrieval for Still Image and Video Databases VII*, 3656, pp. 290-301, 1999.

[9] W.J. Heng, K.N. Ngan, "Post shot boundary detection technique: flashlight scene determination", *Proc. of the Fifth International Symposium on Signal Processing and Its Applications*, pp. 447-450, 1999.

[10] M.-S. Lee, Y.-M. Yang, S.-W. Lee, "Automatic video parsing using shot boundary detection and camera operation analysis", *Pattern Recognition*, Vol. 34, pp.711-719, 2001.