COMBINING LUMINANCE AND EDGE BASED METRICS FOR ROBUST TEMPORAL VIDEO SEGMENTATION

Edmundo Sáez, José I. Benavides

Escuela Politécnica Superior
Dept. Electrotecnia y Electrónica
Avda. Menéndez Pidal s/n
14071 Córdoba, Spain

Nicolás Guil

E.T.S. Ingeniería Informática
Dept. Arquitectura de Computadores
Apdo. Correos 4114
29080 Málaga, Spain

ABSTRACT

This work presents a new video indexing technique able to detect shot cuts as well as gradual transitions in MPEG compressed video sequences. Moreover, this technique also allows to perform global motion estimation in order to successfully detect swing, zoom and displacement effects in video streams. In fact, motion information is used during the gradual shot transition detection algorithm in order to reduce false positives caused by video motion. Two different metrics are used. The first one is based on luminance histograms, while the second is based on the Generalized Hough Transform. The use of this technique greatly improves the results obtained by other shot cut detection algorithms. Also, it is a key technique to the success of both, the gradual shot transition detection and the global motion estimation algorithms.

1. INTRODUCTION

A wide range of video segmentation techniques has been stated [1]. Most of them rely on only one video feature. Nagasaka *et al.* [2] proposed to use luminance histograms. However, spatial information from the images is discarded in the histograms. Also, similarity values are strongly affected by the number of bins in the histograms, as well as gaussian noise. Zabih *et al.* [3] proposed a contour based scene change detection method. However, a registration technique, which is usually very time consuming, was needed. Also, no information was given about the reliability of the computed similarity values when the number of contour points was too low.

Recently, works have focused on scene change detection in MPEG compressed video [4, 5, 6] and frequency domain analysis [7]. This work presents a new video segmentation technique that works with MPEG compressed video using DC images. This technique performs not only shot cut and gradual shot transition detection but also global motion estimation. MPEG motion vectors are not used to perform global motion estimation as they are prone to be erroneous [8]. Two different metrics based on different video features, luminance and contour information, are used. This way, a more reliable description of a video frame is obtained.

The paper is organized as follows. Section 2 describes the metrics actually used. Sections 3, 4 and 5 describe shot cut detection, global motion estimation and gradual shot transition detection respectively. Results are included within these sections. Finally, conclusions are given in Section 6.

This work has been partially supported by the Spanish Government, within the project TIC 2003-06623

2. METRICS FOR VIDEO ANALYSIS

As it is said before, two different metrics are used. Details about these two metrics are given below.

2.1. Luminance based distance function

Luminance histograms are built for both DC images to be compared. Classical histogram comparison techniques perform poorly when applied to histograms built from DC images. Thus, a modified normalized 1D cross correlation process is performed. Its expression is given below:

$$LBDF(H_1, H_2) = \frac{\sum_b H_1(b) \cdot W(H_2(b))}{NT_1 \cdot NT_2} \qquad (1)$$

where

$$NT_x = \sqrt{\sum_b H_x(b) \cdot W(H_x(b))} \qquad (2)$$

$$W(H(i)) = \sum_w H(i+w). \qquad (3)$$

H_1 and H_2 are the luminance histograms, b is a histogram bin, and w is the width of a window centered at bin i.

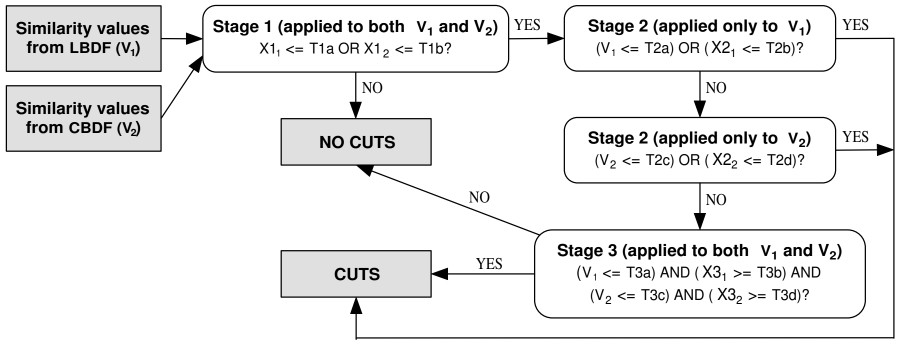


Fig. 1. Thresholding process for the shot cut detection algorithm.

2.2. Contour based distance function

The Generalized Hough Transform (GHT) performs pattern recognition by computing the rotation, scale and displacement of a template in an image. Three transformations are defined. Transformation T is used to compute rotation angle. Scale factor is computed using transformation S. Finally, displacement is computed using transformation D. A complete description of the GHT can be found in [9].

This technique is used as a distance function to obtain a similarity value between two DC images. It also allows swing, zoom and displacement detection in video streams by means of the transformations aforementioned [9].

3. SHOT CUT DETECTION

The shot cut detection algorithm is the first step of the global motion estimation and gradual shot transition detection algorithms, as they analyse video segments between previously detected shot cuts. It uses both metrics defined in Section 2 in order to characterize video frames. Two sets of similarity values, V_1 and V_2, are computed from the video stream using the Luminance (LBDF) and the Contour Based Distance Function (CBDF) respectively. Even though all similarity values in V_1 are supposed to be reliable, the same is not applicable to V_2. In fact, it can be stated that, if the number of contour points in a DC image is lower than p, the similarity value obtained using that DC image is not reliable, as very little information has been used. In order to formally determine a suitable value for parameter p, a divergence estimator has been used [10]. Hence, a suitable value for p is about 70 contour points. The use of the CBDF greatly improves the results obtained when only luminance histograms are used.

Once the similarity values have been computed the thresholding process in Figure 1 is performed. It uses a 21 frame sliding window. In order to characterize a shot cut, three parameters are defined. Parameter X_1 is the value of the distance function in the middle of the observation window. Shot cuts are supposed to exhibit a low value of X_1. Parameter X_2 is X_1/M_w, where M_w is the minimum value of the distance function along the observation window, excluding the central element. X_2 is typically low during a shot cut. Finally, parameter X_3 is M_w and is used to detect masked shot cuts, which are very common in situation comedies. A deep study, based on graphic representation of parameter values, has been carried out in order to find out suitable values for the thresholds in Figure 1, obtaining threshold values that make the method sequence independent. This method performs at real time processing speed. A more detailed description is given in [10]. Obtained results using several videos from the MPEG-7 Content Set [11] are given in Table 1.

Sequence type	Ref	Length	C	M	F	R	P
News	V3	28’ 33”	251	1	17	1.00	0.94
News	V4	18’ 27”	97	0	3	1.00	0.97
Film	V6	15’ 57”	103	9	11	0.92	0.90
Serial	V7	15’ 36”	141	1	2	0.99	0.99
Basket	V16	15’ 38”	76	0	7	1.00	0.92
TOTAL		**94’ 11”**	668	11	40	**0.98**	**0.94**

Table 1. Experimental results for the shot cut detection algorithm (C: Correct; M: Missed; F: False; R: Recall; P: Precision).

4. GLOBAL MOTION ESTIMATION

The basis of the global motion estimation algorithm is to mark every frame in the video stream where a motion effect is suspected to be taking place. A sliding n frame window is used (in this work, $n = 10$), the first frame being the one that is being analysed. The three transformations defined by the GHT (T, S and D) are used in order to detect swing, zoom and displacement effects. The first frame in the window (f) is compared to the next ones

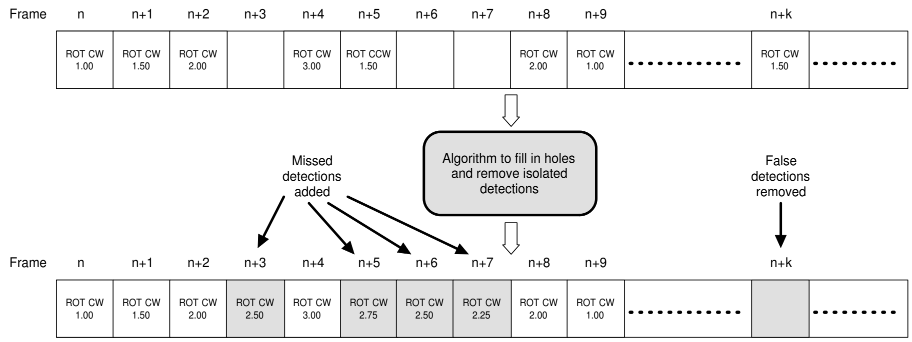


Fig. 2. Operation of the algorithm to fill in holes and remove isolated detections.

$(f+1, f+2, \ldots, f+n-1)$ obtaining $n-1$ values of rotation $(RV(0), \ldots, RV(n-1))$, scale $(SV(0), \ldots, SV(n-1))$ and displacement $(DV(0), \ldots, DV(n-1))$. Further study of these values will allow to mark frame f as a swinging, zooming and/or displacing frame, according to rotation, scale and displacement values respectively.

In order to mark a frame as a swinging frame, rotation values $RV(i)$ are analysed. If most of them are 0, no swing effect is declared. However, if RV shows increasing (decreasing) values, a counterclockwise (clockwise) swing effect is taking place. The magnitude of the values in RV will be useful to estimate swing speed. Zoom detection is very similar to swing detection. However, scale values $SV(i)$ are used. Finally, displacement detection is performed using displacement values $DV(i)$. Displacement angle α_d and relative displacement speed s_d are computed. Pan (P_R) and tilt (T_R) rate in pixels per frame can be directly obtained from α_d and s_d using the following expressions:

$$P_R = 8 \cdot s_d \cdot \cos \alpha_d \quad (4)$$

$$T_R = 8 \cdot s_d \cdot \sin \alpha_d. \quad (5)$$

Lastly, as a swing, zoom or displacement effect lasts for several frames, some of them might remain unmarked. Thus, a process to properly mark these frames has to be carried out. Also, isolated detections lasting for less than a predefined number of frames should be removed. This process is depicted in Figure 2. A complete description of this method is given in [9]. Figure 3 shows displacement angle and pan/tilt rate for the well known sequence *coast guard*. Displacement angles for the three different displacement effects in the sequence are shown in Figure 3(a), while pan and tilt rates are in Figure 3(b).

5. GRADUAL SHOT TRANSITION DETECTION

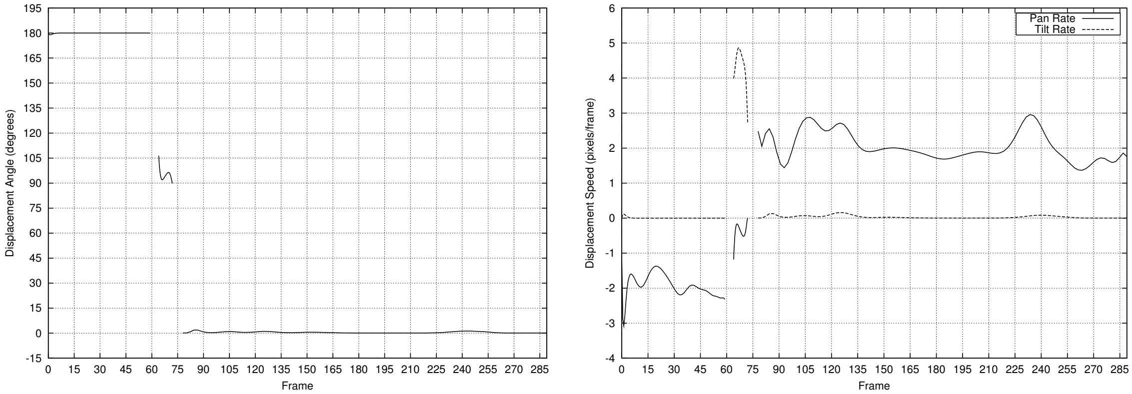
The gradual shot transition detection algorithm uses a similar scheme to that used in the global motion estimation algorithm. A sliding n frame window is used, the first frame being the one that is being analysed. Every frame in the video stream where a gradual shot transition is suspected to be taking place is marked. Two image change rate estimators are defined as follows:

$$ICR_F = \frac{1}{W_L - 1} \cdot \sum_{i=1}^{W_L - 1} \frac{1 - LBDF(H_0, H_i)}{i} \quad (6)$$

$$ICR_B = \frac{1}{W_L - 1} \cdot \sum_{i=0}^{W_L - 2} \frac{1 - LBDF(H_{W_L - 1}, H_i)}{W_L - i - 1}. \quad (7)$$

ICR_F is the forward image change rate estimator. The similarity values between the first frame in the sliding window and the next ones are computed and weighed. The further the two DC images in the window, the lower the weight that is given to the similarity value. Then, the mean value of the weighed similarity values is calculated. ICR_B is backwards image change rate estimator. It is computed using the same process stated for the ICR_F but the last frame of the sliding window is used instead of the first. $W_L = n$, where 10 is a suitable value. Two estimators are used because they both allow accurate detections of the starting and ending frames of a gradual shot transition. The frame under consideration is marked if both estimators exceed a predefined threshold. Finally, a process similar to that depicted in Figure 2 is performed. Gradual shot transition detection algoritmhs usually produce a high amount of false detections due to motion. Thus, the last stage in the algorithm is to perform global motion estimation (see Section 4) in the frame range where a gradual shot transition has been declared, in order to discard false detections.

The method has been tested using almost 20 min. of video sequences including 148 gradual transitions. Two kind of tests have been performed. The first one deals with effects. A gradual shot transition is said to be correctly detected if at least 50% of the frames have been marked. Obtained recall is 0.85 while precision is 0.95. 63% of the false detections were due to motion and removed by the global motion estimation algorithm. The second one deals with



(a) Displacement angle

(b) Pan/tilt rate

Fig. 3. Displacement angle and pan/tilt rate for *coast guard* sequence.

frames and is related to the accuracy of detections. 1824 frames were correctly marked, 250 were missed and 122 were false detections. Thus, accuracy of detections exhibits 0.88 of recall and 0.94 of precision.

6. CONCLUSIONS

A new video indexing technique able to detect shot cuts and gradual shot transitions in video streams has been stated. Moreover, a global motion estimation algorithm has also been developed. It allows the detection of swing, zoom and displacement effects. The use of the GHT is a key element to this video indexing technique. Results clearly surpass those from classical methods, turning this proposal into one of the best ever stated.

7. REFERENCES

[1] I. Koprinska and S. Carrato, "Temporal video segmentation: A survey," *Signal Processing: Image Communication*, vol. 16, no. 5, pp. 477–500, January 2001.

[2] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full-video search for object appearances," in *IFIP Proceedings, Visual Database Systems*. 1992, vol. 2, pp. 113–127, Elsevier Science Publishers B.V. (North-Holland).

[3] R. Zabih, J. Miller, and K. Mai, "Feature-based algorithm for detecting and classifying production effects," *ACM Journal on Multimedia Systems*, vol. 7, no. 2, pp. 119–128, 1999.

[4] B. Yeo and B. Liu, "Rapid scene analysis on compressed videos," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp. 533–544, 1995.

[5] J. Bescós, "Real-time shot change detection over online MPEG-2 video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 4, pp. 475–484, 2004.

[6] D. Lelescu and D. Schonfeld, "Statistical sequential analysis for real-time video scene change detection on compressed multimedia bitstream," *IEEE Transactions on Multimedia*, vol. 5, no. 1, pp. 106–117, 2003.

[7] S. Porter, M. Mirmehdi, and B. Thomas, "Temporal video segmentation and classification of edit effects," *Image and Vision Computing*, vol. 21, no. 13–14, pp. 1097–1106, 2003.

[8] J. Heuer and A. Kaup, "Global motion estimation in image sequences using robust motion vector field segmentation," in *Proc. ACM Multimedia 99*, Orlando, FL, USA, October 1999, pp. 261–264.

[9] E. Sáez, J.M. Palomares, J.I. Benavides, and N. Guil, "Global motion estimation algorithm for video segmentation," in *IS&T/SPIE Symposium Proceedings, Visual Communications and Image Processing (VCIP 2003)*, Lugano, Switzerland, July 2003, vol. 5150, pp. 1540–1550.

[10] E. Sáez, J.I. Benavides, and N. Guil, "Reliable real time scene change detection in MPEG compressed video," in *Proceedings IEEE International Conference on Multimedia and Expo (ICME 2004)*, Taipei, Taiwan, June 2004.

[11] MPEG Requirements Group, "Guide to obtaining the MPEG-7 content set," Doc. ISO/MPEG N2570, December 1998, MPEG Rome.