The SAMOVA Shot Boundary Detection for TRECVID Evaluation 2004

Gaël Jaffré Philippe Joly Siba Haidar

SAMOVA - IRIT
Université Paul Sabatier - 118 route de Narbonne
Toulouse, France

Abstract

This paper describes the system used by the SAMOVA team (IRIT) to perform the shot boundary detection task of the TRECVID 2004 workshop. Our main algorithm is based on detection of illumination changes, as well as on the length of these transitions. The results are then refined by ckecking all the candidate transitions.

We propose three versions of our application.
Samova01: the simplest form, only uses illumination changes with image difference.
Samova02: checks every cut, using morphological information.
Samova03: checks both cuts and gradual transitions.
The second and third runs are improved versions of the first run. They use a small threshold to have more detections, but add a validation test for each detected transition.

Our application is made to work with compositing effects, which are editing effects. So, the evaluation of our application makes sense if we take into consideration compositing effects.On video contents where such effects are representative of the editing work our results are improved by the ability of our methods to deal with effects such as overlay, captions suddenly appearing on the screen, ...

1 Introduction

The SAMOVA team from IRIT has this year its first participation to the TRECVID evaluation workshop, with three submissions for the shot boundary detection task. The submitted algorithms rely on a statistical analysis of the intensity pixel variation along the time in presence or not of transition effects. This approach was published in 1994 [1] and upgraded in 2000 [6] to also detect compositing effects.

2 Shot Boundary task

Segmentation into shots is one of the basic process that must be applied to edited videos so as to extract information on the document structure. This information is useful for post-processing operations on video content. A clear definition of a shot can be found in [6]. There are many transitions effects, but for TRECVID we consider that every transition which is not a cut is called gradual transition.

We present in this paper three variants of our algorithm of shot boundary detection. The first one only uses illumination changes, and will be described in section 2.1. The second and third ones uses the main detection with a weak threshold, but then validate each candidate with morphological information. There are presented in section 2.2.

Figure 1: Each frame is resized in order to decrease processing time.

2.1 Transition detection

Our application of transition detection is based upon the estimation of motion. For each frame, its luminance is computed. Then, the dominant colors are extracted, used to estimate the contrast, and then the motion. First, the cut detection is carried out. If no cut is detected, then we run the gradual transition detection.

2.1.1 Pre-processing

To optimize the computational time, we create small images from the original frames, by taking one pixel every eight pixels: it reduces by 64 the size of the frames (example is given in Figure 1). In the TRECVID corpus, the frames have a size of 352×240: thus we obtain a new frame size of 44×30.

Our images extracted from the video streams are coded using the RGB color system. However, to easily extract luminance and dominant colors, we convert each frame into the HSV color space [2], and we only keep the V value for luminance processing.

2.1.2 Cut detection

For each new frame, the absolute difference with the previous frame is computed. Then we count the number of pixels which are different of more than 128 greylevels. If the value is a lot greater than all the values since the last shot, then a cut is detected.

2.1.3 Gradual transitions

In gradual transitions, we only detect dissolves and fades. As gradual transition detection is less reliable than cut detection, cut detection has priority over gradual transitions.

To detect gradual transition, we consider that pixel intensity has the same variation sign on the whole transition. For each pixel of the small frame, we store the direction of the intensity variation. For each new frame, we count the number of pixels that keep the same direction. If it is above a threshold, then we detect the beginning of a gradual transition. We suppose that the transition is over when this number becomes lower than another threshold, as shown in Figure 2.

2.2 Transition validation

We start from the original detection (cf sections 2.1.2 and 2.1.3), with smaller thresholds, which implies more detections. Then detected transitions are not automatically kept, they have to be validated.

To detect a transition, we take the first and the last frames of the candidate, compute the absolute difference, and then we binarize this new image ($0 \Rightarrow$ no change, $1 \Rightarrow$ change). The goal is then to find the connected areas which really changed, without noise. In order to reduce the noise, we use morphological information [3, p. 489]: we apply a dilation with a 7×7 mask, examples are presented in Figure 3 and Figure 4. Then, a candidate is validated if its number of pixels which have changed after the dilation is above a threshold (we take roughly half of the whole frame).

This method corresponds to a fast implementation of compositing effect detection which is much more refined in its original version [7]. This version is included in the MPEG7

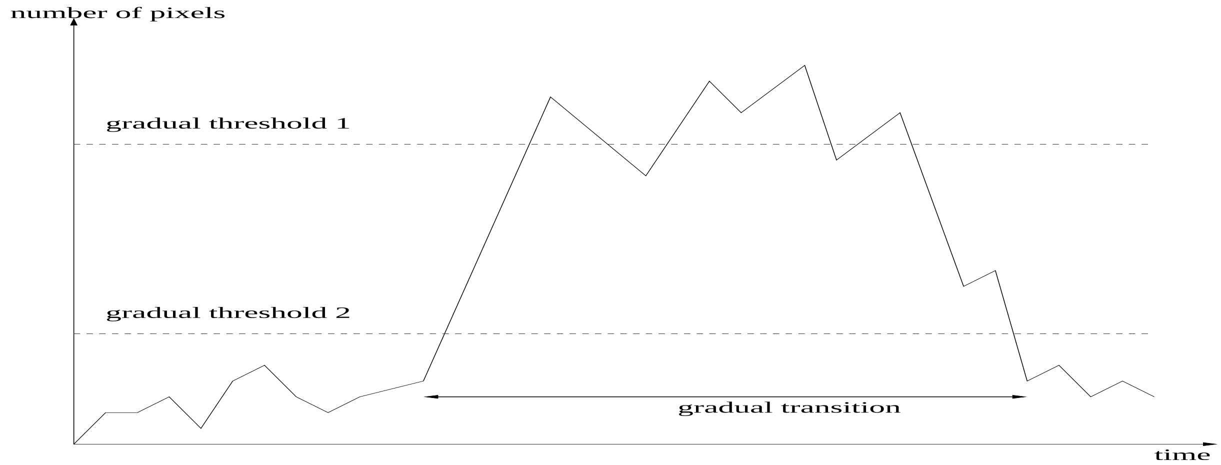


Figure 2: Detection of gradual transitions.

reference software [5] to instantiate the editing work description scheme.

3 Results

For TRECVID 2004, we proposed three runs. In the first one (Samova01), we only applied the detection of cuts and gradual transitions, without validation. For the second one (Samova02), we ran the cut detection with a lower threshold, and then add a test to validate each candidate, whereas the gradual transition detection remains the same. Finally, for the third one (Samova03) we add a test for both cuts and gradual transitions. Results for precision and recall are summarized in Table 1, results about computational time are given in Table 2, and comparison with the other groups are given in Figure 5.

We can note that processing time is very weak, it takes approximatively 90 minutes (1.5 hour) to process 6 hours of videos.

4 Conclusion

We have presented the application used by the SAMOVA Team from IRIT to perform the shot boundary detection task of the TRECVID 2004 workshop. We proposed three runs: the first one (Samova01) is the simplest form, only using illumination changes with image difference. The second one (Samova02) checks every cut, using morphological information. Finally, the third one (Samova03) checks both cuts and gradual transitions.

Our application is made to work with compositing effects (they are integrated in MPEG7 [4]), which are editing effects [7]. So, the evaluation of our application makes sense if we take into consideration compositing effects. Whatever it allows a really fast detection of classical transition effects as a preprocessing step in a deeper video analysis process.

In order to measure the evolution of all the techniques in shot boundary detection, we think that future evaluations should take into account those effects.

Acknowledgments

This work has been conducted on the behalf of the CNRS Pidot project.

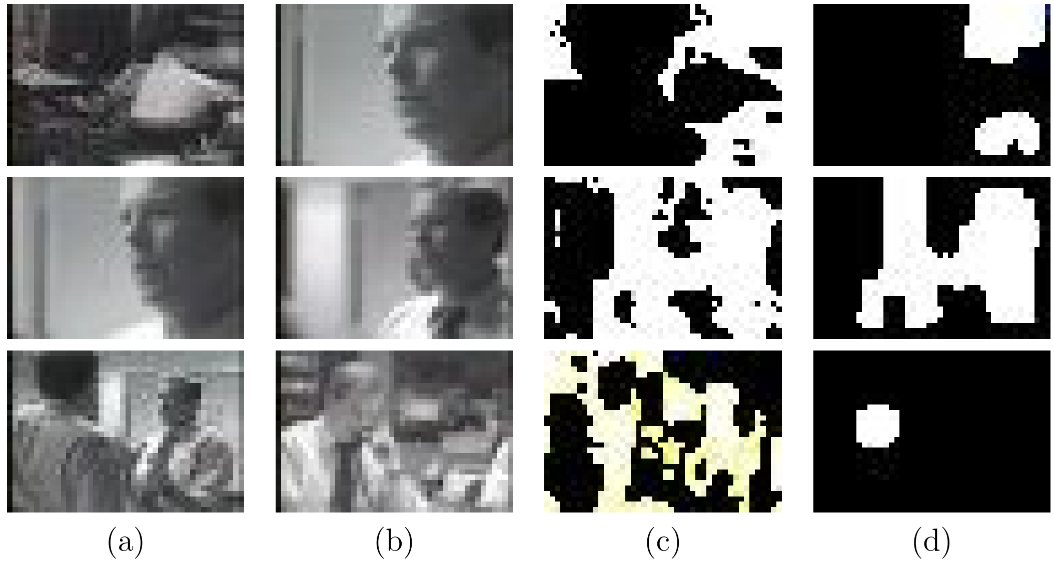


Figure 3: Validation of three candidate cuts. For each candidate, the columns (a) and (b) represent the frames between and after the cut. Black pixels in column (c) are pixels which drastically changed from (a) to (b). Column (d) is the result of dilation of (c). The first and third rows are validated candidates, whereas the second one is a rejected candidate.

Figure 4: Validation of a candidate for gradual transitions.

References

[1] P. Aigrain and P. Joly, "The Automatic Real-Time Analysis of Film Editing and Transition Effects and its Applications," *Computers and Graphics*, vol. 18, no. 1, pp. 93–103, Jan. 1994.

[2] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice-Hall, 2003.

[3] B. Jähne, *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*, fourth ed. Springer, 1997.

[4] B. S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG 7: Multimedia Content Description Language*. Wiley-Interscience, 2002.

[5] MPEG7 Reference Software, www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html

[6] R. Ruiloba and P. Joly, "Framework for Evaluation of Video-to-Shots Segmentation Algorithms," *Networking and Information Systems Journal*, vol. 3, no. 1, pp. 77–103, 2000.

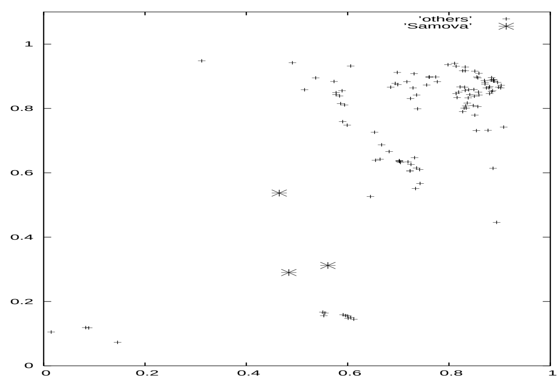
[7] R. Ruiloba and P. Joly, "Adding the concept of video editing levels in shot segmentation," in *Proceedings of the 9th Information Processing and Management of Uncertainty in Knowledge-based Systems*, vol. 1, Annecy, France, July 2002, pp. 501–506.

	All		Cut		Gradual Transition			
							Frame	
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
Samova01	0.561	0.312	0.812	0.372	0.030	0.031	0.175	0.679
Samova02	0.484	0.290	0.698	0.351	0.030	0.031	0.176	0.689
Samova03	0.465	0.537	0.678	0.584	0.014	0.059	0.195	0.683

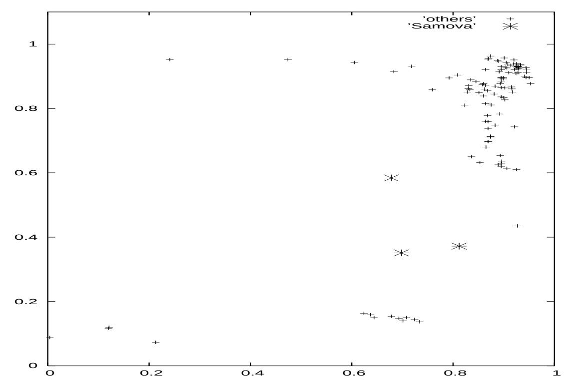
Table 1: Results for the Samova runs.

	run time	decode time	segmentation time
Samova01	5389	2711	2678
Samova02	5405	2711	2727
Samova03	5387	2711	2676

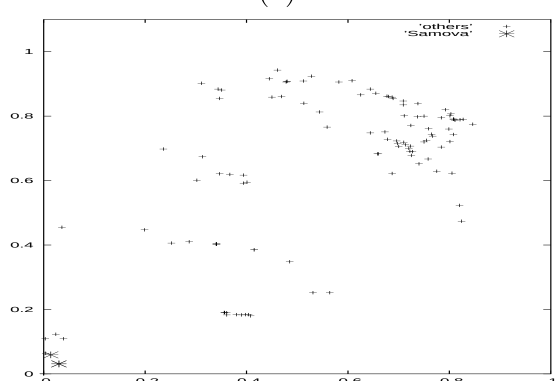
Table 2: Computational time for the Samova runs. Times are given in seconds, and the runs were processed on a Pentium IV 1.7Ghz, with a C implementation.



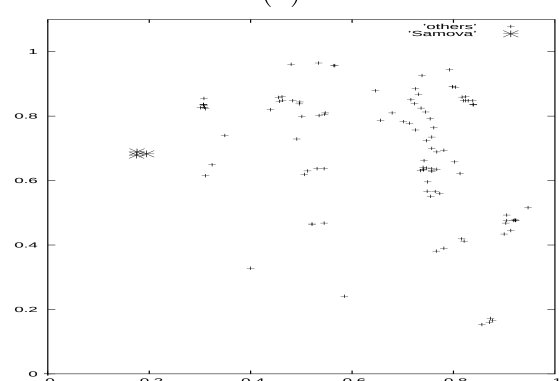
(a)



(b)



(c)



(d)

Figure 5: Recall×Precision. (a) for all transitions. (b) for cuts. (c) for gradual transitions. (d) Frame-recall×Frame-precision global results for gradual transitions.