World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:4, 2007

Spatio-Temporal Video Slice Edges Analysis for Shot Transition Detection and Classification

Aissa Saoudi, and Hassane Essafi

***Abstract*—**In this work we will present a new approach for shot transition auto-detection.

Our approach is based on the analysis of Spatio-Temporal Video Slice (STVS) edges extracted from videos. The proposed approach is capable to efficiently detect both abrupt shot transitions "cuts" and gradual ones such as fade-in, fade-out and dissolve.

Compared to other techniques, our method is distinguished by its high level of precision and speed.

Those performances are obtained due to minimizing the problem of the boundary shot detection to a simple 2D image partitioning problem.

***Keywords*—**Boundary shot detection, Shot transition detection, Video analysis, Video indexing.

I. Introduction

WITH the recent advances in hardware and software areas, many facilities have been produced for using and sharing digital multimedia documents such as images, audios and videos.

In many cases, those facilities have led, by ignorance or with intention to much illegal exploitation. This attitude is increased with the widespread of the peer-to-peer network and recently the UGC (user-generated content) solutions.

In order to detect the presence of an illegal copy, some emerging solutions propose the identification of similar multimedia documents. There are two approaches used for copy detection: the watermarking based approach and signature based approach.

In our company we are interested in the signature-based approach, in which the extracted signatures characterize the contents of the analysed document. The main tasks required for signature extraction are boundary shot detection and shot description.

In this paper our work is focused on boundary shot (shot transition) detection. Boundary shot detection plays an important role in many applications such as the management of multimedia databases, the creation of automatic summaries of videos and televised series, the compression of video sequences and tracking objects in real time in dynamic scenes.

Manuscript received May 25, 2007.

Saoudi Aissa is with University Paris 13 (Institue Galilée) and R&D Department, Advestigo Compagny, 140 Bureaux de la Collines 92210 Saint-Cloud, France (corresponding author to provide phone: +33 1 72 77 70 17; fax: +33 1 46 89 68 60; e-mail: aissa.saoudi@advestigo.com).

Hassane Essafi. R&D Department, Advestigo Compagny, 140 Bureaux de la Collines 92210 Saint-Cloud, France (phone: +33 1 72 77 70 03; fax: +33 1 46 89 68 60; e-mail: hassane.essafi@advestigo.com).

Our main motivation is shot localisation. The localised shots will then be characterised and eventually used for video characterisation. This strategy allows a video local characterization which gives the possibility to conduct partial researches in our characterized video.

The remainder of this paper is organised as follows: in section II and III we will give some basic definition about video structure and the various shot transition types. Section IV consists of previous works on boundary shot detection. Section V describes the method we propose for boundary shot detection and classification. Section VI contains the experimental results obtained and finally, section VII concludes the paper.

II. A Video Hierarchy

A video can be represented in hierarchal manner. In the low level we find all the frames of the video. Frames taken with the same source (generally a camera) without interruption are grouped to form what is called a ***shot***. The continuity of the spatio-temporal and visual information in those frames is one of the main features of a shot.

Successive shots that transport certain complete semantic information can be grouped to form what is known as a scene (see Fig. 1) to simplify the meaning, a video can be considered as an analogue of a text. Frames, shots and scenes in a video can be considered as letters, words and sentences in a text respectively.

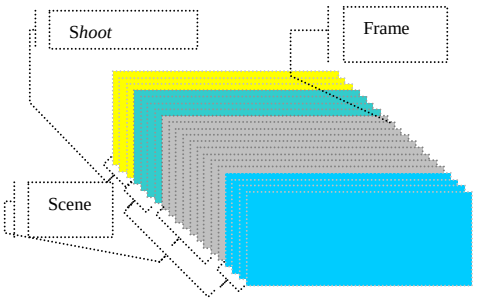


Fig. 1 A video hierarchy

III. Shot Transition Types

The shots are assembled during the editing phase using a variety of techniques which are more or less complex and producing different kinds of shot transitions: *Cut*, *Fade-in*, *Fade-out* and *Dissolve*.

Cut is the abrupt change in scene or camera shot that comes from editing two different shots together.

Fade, Dissolve: all of these editing techniques are slow transitions between shots, less abrupt than a cut. A *fade* comes in two varieties: a *fade-in* and a *fade-out*. A fade-in starts from a solid colour screen (usually black, but sometimes white and rarely other colours) and slowly transitions to a shot in the movie, as the shot is superimposed over the solid screen. A fade-out starts with a shot and transitions to a solid colour. A *dissolve* is similar to a fade, but instead of moving between a shot and a solid colour, it moves between two images.

IV. Related Work

Several approaches have been proposed for boundary shot detection based on the spatio-temporal information continuity between frames of the same shot. It can be classified into many types amongst which are the most popular ones including Pixel Differences, Histogram and Statistical differences, Edge differences, motion based approaches and Shot transition detection in the compressed domain .

- ***Pixel Differences Based Approaches***

The easiest way to compare tow frames is to compare their colorimetric pixels values one by one.

T. Kikukawa and *S. Kawafuchi* accumulated the differences in pixels intensities between two consecutive frames. Then, this accumulation is compared to a predefined threshold to detect the presence of a shot transition [1]. *H. Zhang* et al proposed the use of a method called 'pair-wise pixel comparison' which is based on the use of a predefined threshold to determine the percentage of pixels that are considerably changed. This percentage is compared to a second threshold to detect a shot transition[2].

The advantage of those approaches is the simplicity of their implementation; however they are sensitive to scene objects motions, zooms and movements of the camera. As a result many false alarms will be triggered.

- ***Histogram and Statistical Differences Based Approaches***

The most used approaches are those based on histograms and their variations [3-6]. According to *Ueda* et *al* the rate of changes in colour histogram which correspond to the successive frames, was determined in order to detect the presence or the absence of a shot transition. *A. Nagasaka* and *Y. Tanaka* proposed the comparison of the grey level histograms between two frames. The histogram $H_n(k)$ is obtained by counting the number of pixels in a frame witch have the same grey level *k*. The difference between two histograms is then determined using:

$$D_n(H_{n-1},\ H_n) = \sum_{k=1}^{K} \left| H_{n-1}(k) - H_n(k) \right|$$

Where *K* is the number of grey levels. If D_n is greater than a given threshold a shot cut is declared. *Ishwar K. Sethi* and *Nilesh* Patel extracted the *I* frames from the Mpeg compressed video and then they computed a histogram for each frame. In order to detect the presence of cuts, they obtained good results by applying the following statistic tests (*Yakimovsky* Likelihood ratio test, Chi-Square test and *Kolmogorove-Smirove* test)[3].

In general, histogram-based approaches are characterised by a weak detection rate because the spatial distribution of information is disregarded.

- ***Edge Differences Based Approaches***

Zabih et *al* based their studies on the idea that edges of the last frame in one shot are different from those of the first frame in the next shot[4].

They proposed the use of the correlation between edge pixels of two consecutive frames. This approach is known under the mane of "edge change ratio".

The drawback of those kinds of approaches is their relatively high number of false alarms, especially in scenes with high speed motion such as action movies and music video clips.

- ***Motion Based Approaches***

In previous approaches, many cases of false alarms are caused by camera movement and scene objects motion. So, some works have been proposed to detect motion. This detection can be used in two different ways. The first, the motion break can indicate a shot transition. Motions in the second one can be detected and eliminated in order to suppress their impact on the difference between consecutive frames. This task improves the results obtained in other approaches due to a decrease in the number of false alarms caused by a motion.

Recent works were focused on the use of block displacement for the detection of motion vectors. This is achieved by searching for each block in frame f_n the best matching block in a neighbourhood around the corresponding block in frame f_{n+1}. This principle is shared between several proposed approaches. However, the main difference between them is in the way of comparing blocks and in resuming the similarities between consecutive frames in only one value.

Akutsu et *al* have used the average of the maximum correlations between every two blocks[5]. Generally, the combination of poor matches with good ones gives a passable match between two frames belonging to the same shot. To avoid this, *Shahraray* proposed the use of a filter in which some worse matching blocks can be discarded[6]. But this can reduce the rate of shot transition detection. To overcome this problem, they suggested that a certain number of good matching blocks should also be excluded.

S.Porter proposed the comparison of blocks using the correlation between block edges[7]. This is reinforced by measuring the distance between colour histograms. Similar block localisation in consecutive frames allows an estimation of the continuity of the motion, and hence detection of the continuity of the shot. In the absence of the continuity of the motion a shot transition will be detected.

Motion based approaches improve the rate of boundary shot detection, but their high computational cost makes them unlikable.

- ***Shot Transition Detection in the Compressed Domain***

To discard the video decompression task, in some proposed approaches, the detection is made directly on the compressed data[8, 9].

De Bruyne et *al* proposed the detection of shot transition in H.264/AVC compressed video. To achieve this, they first analysed the temporal predictions used in the *P* and *B* frames.

For two successive frames, if the main prediction of the first one is a temporal backward and a temporal forward for the next one, then a shot transition will be signalled. For the *I* frames, the detection is made by the analysis of the frequency of the 4x4_MODE and 16x16_MODE used. If the difference between these frequencies is significant, it indicates that those frames are contained in two different shots [10].

Discarding the decompression task reduces the processing time; however this requires a number of adaptive algorithms for every compressed type.

V. The Proposed Algorithm

Our application concerns video copy detection, in which we need to be able to detect a video sequence even if alternated (such as resizing, shifting, camcording, etc.) or included in a long sequence. This can not be done without having a boundary shot detection (shot transition detection) robust against usual video manipulations. We have evaluated some of the existing boundary detection methods but the results were not very satisfying for our application.

In this paper we propose a new approach based on the analysis of Spatio-Temporal Video Slice (STVS) edges which are extracted from videos and that we call STVSE. Before going into the details of the proposed method, we first describe the STVS concept.

A. Spatio-Temporal Video Slice Concept

A video can be considered as a parallelepiped, its volume represents the colorimetric values of pixels in consecutive frames. By slicing this parallelepiped through the time axis, we obtain two parts where the new faces represent an STVS (shown in Fig. 2).

Concretely, the STVS can be built by selecting one segment of pixels (L_t) from each frame and concatenating them to formulate a 2D image.

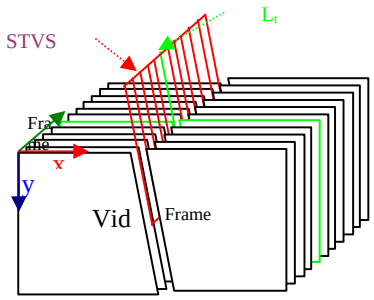


Fig. 2 STVS Extraction

This concept was initially proposed by *Edward H Adelson* and *James R Bergen* [11] for the computational vision tasks.

Despite the huge importance of the information contained in the STVS, very few works was dedicated to its analysis and investigation in order to create better solutions.

B. STVS Edges Analysis

The STVS (Fig. 3 (a)) contains a huge quantity of information about objects motion, cameras zooms and motions including tilts and pans. Moreover, it contains important information regarding shot transitions (cuts, fades, dissolves, etc.) which highlights our main target.

By analysing the STVS, it was noticeable that shot transitions split the STVS, in the direction of the time axis, into two homogeneous zones. This is due to the disappearance of scene objects/zones of the current shot and the appearance of the components of the next. Therefore, to detect the shot transition, we can analyse the continuity of the STVS signal in the time axes. This can be done by processing and analysing the image edges of the STVS image. As can be seen from Fig. 3 (b) which represents the produced image edges, the boundary shots correspond to the perpendicular segments against the time axis in the image edges.

The image edges is obtained via the computation of the STVS gradient magnitude image using the *Deriche* detector[13] . Then by applying the *Savola* algorithm [14], the image is binarised. Fig. 3 (a) and Fig. 3 (b) show the STVS and STVS Edges of “Bug’s life” film, successively.

Fig. 3 (a) STVS of the first 200" "Bugs life"

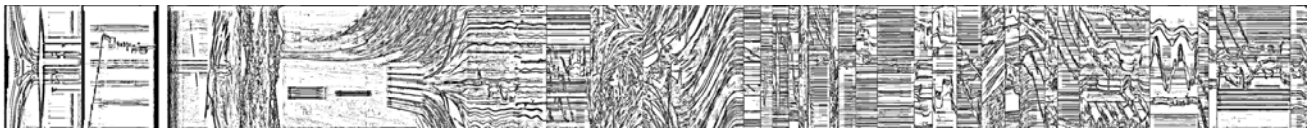


Fig. 3 (b) STVS image edges of the first 200" "Bugs life"

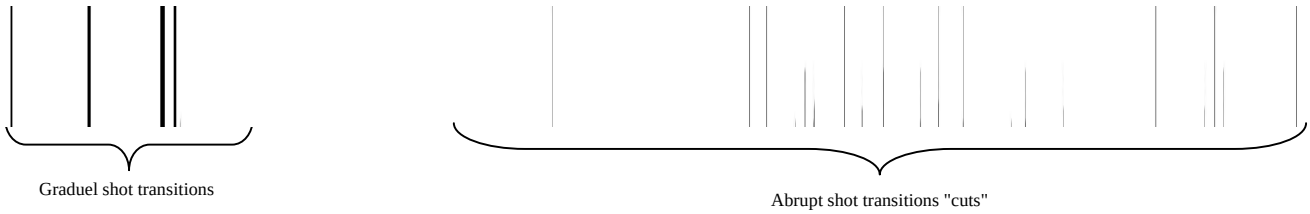


Fig. 3 (c) Detected shot transitions

C. Detection of Shot Transition

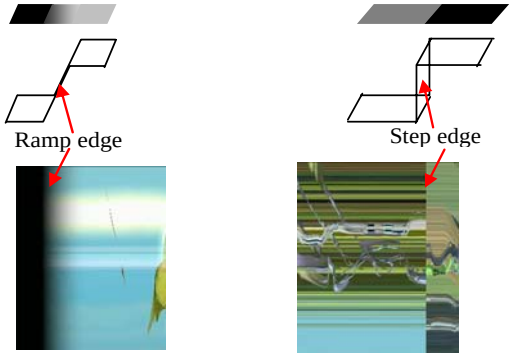
As can be seen from Fig. 3 (b), the shot transition corresponds to the perpendicular segments of binary image STVS edges against the time axis. To detect those segments, we first produce a binary image histogram in which each bin corresponds to an image edges column, and its amplitude represents the number of perpendicular segments (continuous edge pixels) to the time axis. Bins with amplitudes above a given threshold (dominants bins) correspond to boundary shot positions (see Fig. 3 (c)).

D. Shot Transitions Classification

After detecting the shot transition, we will now try to determine their types by studying their effect on the histogram shape. We notice that there are some relationships between the distribution of the selected bins and the type of the shot transition:

- Abrupt shot transitions ("cut": Fig. 4 (b)) correspond to abrupt discontinuities in the content of the STVS on the time axis, which consequently produce isolated dominant bins in the histogram (Fig. 4 (d)).
- Graduate shot transitions are characterized by non-isolated dominant bins, due to the fact that the content varies gradually. (Fig.4 (a)-(c)).

Furthermore gradual transitions can be classified into three main classes: fade-in, fade-out, and dissolve. To distinguish between these three types, we analyze the left and right regions of each gradual shot transition in the STVS. If the left region is constituted of a homogenous colour, this means that the shot transition is fade-in. In case the homogenous colour is present in the right region; we have fade-out transition. Finally, if both regions are not homogenous, we have a dissolve transition.



a) Gradual shot transition(Fade-in) corresponding edge type

b) abrupt shot transition(cut) corresponding edge type

c) Gradual transition corresponding bins

d) Abrupt transition corresponding bins

Fig. 4 Gradual and abrupt shot transition corresponding edge type and bins

VI. Experiments and Results

The main objective of our experiment is to measure the efficiency of the proposed approach, to detect and to classify shot transitions in a given video sequences. We use the classical definition of recall and precision.

$$\mathrm{Re}\,call = \frac{True\ Positives}{True\ Positives + False\ negatives}$$

$$\Pr ecision = \frac{True\ Positives}{True\ Positives + False\ positives}$$

In this experiment, a variety of videos were utilised including movies, cartoons, sports, news and commercial spots. Their duration vary between few seconds and twenty minutes.

We have calculated the number of shot transitions for each type. These are summarised in Table I.

TABLE I
THE NUMBER OF SHOT TRANSITIONS PER TYPE

Transition type	*Cut*	*Fade-in*	*Fade-out*	*Dissolve*
Number	520	7	6	3

Table I and Fig. 5 indicate that abrupt shot transition is the main technique used in the editing of shots.

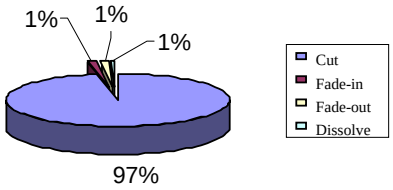


Fig. 5 The percentage of shot transition utilization by type in the experimental video sequences

Table II gives the number, the precision and recall for each type of shot transition detection.

TABLE II
THE RESULTS OF SHOT TRANSITION DETECTION

	Cut	*Fade-in*	*Fade-out*	*dissolve*	**Total**
Total number of transitions	*520*	*7*	*6*	*3*	**536**
Detected transitions	*464*	*7*	*6*	*3*	**480**
Missed alarms	*56*	*0*	*0*	*0*	**56**
False alarms	*28*	*0*	*0*	*0*	**28**
Recall	*%89,2*	*%100*	*%100*	*%100*	**%89,5**
Precision	*%94,3*	*%100*	*%100*	*%100*	**%94,4**

The total recall and precision values obtained were %89.5 and %94.4 respectively. According to the results obtained, the algorithm is characterised by high precision and recall. This shows its good performance for detecting both gradual and abrupt shot transitions.

Table III summarizes the results of the evaluation of shot transition classification. The test is based on calculating the number of both correct and wrong classifications.

TABLE III
THE RESULTS OF SHOT TRANSITION DETECTION

	Cut	*Fade-in*	*Fade-out*	*dissolve*	**Total**
Total number	*520*	*8*	*6*	*3*	**537**
correct classified	*448*	*6*	*5*	*3*	**462**
Number of missed classification	*72*	*2*	*0*	*0*	**74**
Number of False classification	*28*	*2*	*2*	*3*	**35**
Recall	*%86,2*	*%75,0*	*%100*	*%100*	**%86,2**
Precision	*%94,1*	*%75,0*	*%71,4*	*%50*	**%93,0**

The recall is less good because of some wrong classifications, where some cuts were considered as gradual transitions. This has badly affected the precision in the classification of gradual shot transitions. This impact is very significant due to the lower number of gradual transitions compared to cut ones.

VII. CONCLUSION

In this paper, we propose a new method for the auto-detection and classification of shot transitions in a video flux.

Our approach is based on the analysis of Spatio-Temporal Video Slice (STVS) edges. They provide important information on motions and spatial object distributions. Additionally, they contain other information with regards to the abrupt and gradual shot transitions.

In our method, the detection of shot transitions depends on the detection of vertical segments that are present in the STVS edges. These segments are produced by joining the end of the homogenous regions in one shot to the start of those in the subsequent shot.

The proposed detector is commonly used to detect the majority of the existing shot transitions, and to automatically classify them into the following types: cut fade-in, fade-out and dissolve.

REFERENCES

[1] T. Kikukawa, S.K., *Development of an automatic summary editing system for the audio-visual resources.* Transactions on Electronics and Information, 1992: p. J75-A(2):204–212,

[2] HongJiang Zhang, A.K., Stephen W. Smoliar *Automatic partitioning of full-motion video.* 1993. **1**(1): p. 10-28

[3] Ishwar K Sethi, N.P., *A statistical approach to scene Change Detection.* 1995. **Vol. 2420**.

[4] Ramin Zabih , J.M., Kevin Mai *A feature-based algorithm for detecting and classifying production effects.* Multimedia Systems, 1999. **7**(2): p. 119 - 128.T

[5] A. Akutsu, Y.T., H. Hashimoto, and Y. Ohba, *Video indexing using motion vectors.* SPIE Visual Communication and Image Processing, 1992. **1818**: p. 1522–1530.

[6] Shahraray., B., *Scene change detection and content-based sampling of video sequences.* SPIE Conference on Digital Video Compression: Algorithms and Technologies, 1995. **2419**(2-13).

[7] Sarah, *Segmentation and Indexing using Motion Estimation*. 2004, University of Bristol.

[8] Chou, S.-C.P.Y.-Z., *Effective wipe detection in MPEG compressed video using macro block type information.* Multimedia, IEEE Transactions on, 2002. **4**(3).
[9] Fernando W.A.C, L.K.K., *Abrupt and gradual scene transition detection in MPEG-4 compressed video sequences using texture and macroblock information.* Image Processing, 2004. **3**: p. 1589-1592.
[10] De Bruyne Sarah, D.N.W., De Wolf Koen, De Schrijver Davy, Verhoeve Piet, Van de Walle Rik, *Temporal Video Segmentation on H.264/AVC Compressed Bitstreams.* Proceedings of the 13th International Multimedia Modeling Conference, LNCS 4351 Advances in Multimedia Modeling., 2007. **1**: p. 1-12.T
[11] Edward H. Adelson, J.R.B., *Spatiotemporal energy models for the perception of motion.* Journal of optical America, 1985. **2 No 2**: p. 284-299.
[12] C. W. Ngo, T.C.P., R. T. Chin, *Camera Breaks Detection by Partitioning of 2D Spatio-temporal Images in MPEG Domain.* IEEE Multimedia System 1999. **1**.T
[13] Deriche, R., *Using Canny's criteria to derive an optimal edge detector recursively implemented.* Internat J Comp Vision,1(2), 1987: p. 167-187.
[14] J. Sauvola, M.P.i., *Adaptive document image binarization.* Pattern Recognition, 2000. **33** (2000): p. 225-236.