Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Int. J. Electron. Commun. (AEÜ) 61 (2007) 463 – 468
International Journal of Electronics and Communications AEÜ
www.elsevier.de/aeue
LETTER

Shot-cut detection for B&W archive films using best-fitting kernel

Oguzhan Urhan*, M. Kemal Gullu, Sarp Erturk

Department of Electronics and Telecommunications Engineering, Kocaeli University Laboratory of Image and Signal Processing (KULIS), Veziroglu Campus, Kocaeli University, 41040 Kocaeli, Turkey

Received 25 January 2006

Abstract

A novel shot-cut detection method for black and white (B&W) archive films, employing the best-fitting kernel between two consecutive image frames is proposed in this paper. The best-fitting kernel, obtained using least squares, is used to estimate the second frame of each successive frame pair. The absolute frame difference between estimated and original image frames is utilized as criterion to decide a candidate scene-cut. A double thresholding approach, consisting of global and local adaptive thresholds, is used in this stage. Obtained candidates are forwarded to a confirmation stage to exclude single-frame defects, frequently encountered in archive film sequences. Experimental results show that the proposed method gives favorable result for B&W degraded archive films, compared to methods proposed in the literature.

Keywords: Shot-cut detection; Video restoration; Archive film; Best-fitting kernel

1. Introduction

Shot-cut detection is an important element and generally the first step of video segmentation. Video retrieval, indexing, analysis, semantic description and compression are some prominent applications in which shot-cut detection. Developing shot-boundary detection methods is an active and significant research area as it is almost essential for all video segmentation procedures. Video restoration is an example application of shot-boundary detection in video processing. Restoration of archive video generally requires the segmentation of video as the first step. However, visual degradations in this kind of video make it difficult to achieve an acceptable detection performance. Therefore conventional shot-cut detection algorithms developed for recently captured video sequences commonly fail. We propose a novel scene-cut detection method which usually can deal with degradations and provide good detection accuracy even for seriously degraded archive film sequences.

A video sequence typically consists of camera shots (or scenes) concatenated employing postproduction techniques. A shot is typically composed of an uninterrupted sequence of frames adjacently captured by the same camera. Hard-cuts and gradual transitions are two main types of transitions between two shots. Numerous methods have been proposed for the detection of these kind of scene transitions some of which are reported in [1–8]. These methods can mainly be classified according to their basic characteristic into pixel-based, histogram-based, feature-based and motion-based techniques [1]. A brief literature review is given initially.

A pixel-based method, which mainly considers pixel-wise differences of consecutive frames, is proposed in [2]; where the absolute sum of frame differences is used to detect a scene-cut. The grayscale histogram difference of successive frames is considered in [3] as a metric for scene-cut detection. Color histograms have also been utilized in several color spaces such as RGB, HSV, YUV, Lab, and Luv

* Corresponding author. Tel.: +90 262 335 11 68; fax: +90 262 335 28 12.
E-mail address: sertur@kou.edu.tr (O. Urhan).

1434-8411/$ - see front matter © 2006 Elsevier GmbH. All rights reserved.
doi:10.1016/j.aeue.2006.08.002

as different metrics for this purpose in [4,5]. A feature-based method using edge information is proposed in [6]. This method considers successive frames and uses normalized proportions of entering edges and exiting edges to decide a scene-cut. The method proposed in [7] utilizes a frequency-domain correlation feature to detect scene-cuts. This technique uses overlapping blocks and evaluates the phase-correlation between each co-paired block in successive frames to decide a scene cut in the case of low correlation. Multi-resolution motion estimation for global motion compensation, employing a two-dimensional affine model is utilized as a first step in [8]; then average pixel differences between motion compensated consecutive frames are thresholded using an adaptive threshold to detect scene-cuts.

This paper proposes a novel shot-cut detection technique that is based on the best-fitting kernel between two successive image frames, obtained in least-squares (LS) sense. The best-fitting kernel is employed to reconstruct (i.e. estimate) the second frame from the first frame of each pair. The absolute frame difference between the estimated and original image frame is evaluated as a decision metric. Global and local adaptive thresholds are employed together to decide candidate scene-cuts. A verification stage is also utilized to avoid possible incorrect scene-cut decisions, in case of visual degradations commonly encountered in archive film sequences. The performance of the proposed shot-cut detection method is compared to a large variety of methods (i.e. histogram-based [3], feature-based [6,7] and motion-based techniques [8]). Experimental results show that the proposed method outperforms the compared methods proposed in the literature.

Cut detection has been studied quite well in the literature. However, proposed techniques in the literature mainly focus on non-degraded and color image sequences. Note that this is not the situation considered in this paper that addresses hard-cuts in black and white archive films which commonly have various visual degradations. Detection of hard-cuts is more important for archive film restoration compared to gradual transitions. Usually, it is possible to continue running restoration algorithms through a gradual change without any problems; however, restoration algorithm parameters need typically to be reset after a hard cut for reasonable performance. Therefore the proposed approach is aimed at hard-cuts only and is obviously not suited to detect gradual shot changes.

2. The proposed method

The main idea of the proposed method is that scene-changes can be detected using the best-fitting kernel between consecutive image frames. The best-fitting kernel is used to estimate the second frame of each consecutive image pair from the first frame. If the estimation error is relatively low the consecutive image pair is decided to belong to the same scene. It can be considered as a scene-cut otherwise. The block diagram of the proposed system is given in Fig. 1.

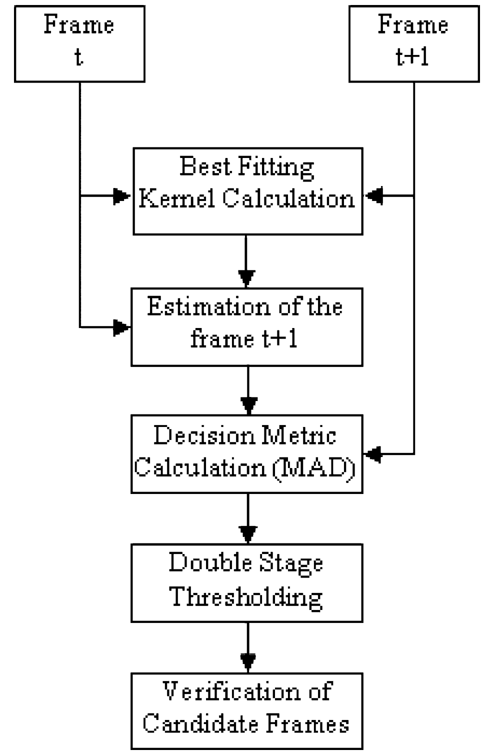

Fig. 1. Block diagram of the proposed system.

2.1. Best-fitting kernel in least-squares sense

Kernels are mainly used for filtering operations in image processing. In this work, the best-fitting kernel for consecutive image frames is calculated in LS sense. If we define a filter kernel $g(p,q)$ of dimensions $l \times l$, then it is possible to obtain an estimate for the next frame by applying a reasonable filtering to the previous frame in the form of

$$\widehat{I}(i,j,t+1) = I(i,j,t) * g, \qquad (1)$$

where $\widehat{I}(i,j,t+1)$ represents the estimate of frame $t+1$ obtained from frame $I(i,j,t)$.

In order to obtain the best possible reconstruction, the filter kernel g is obtained as the best-fitting kernel in LS sense. For this purpose, the optimal (in LS sense) kernel is calculated by solving

$$\min_{g(p,q)} \|I(i,j,t+1) - I(i,j,t) * g\|_2^2. \qquad (2)$$

Obviously, the estimation accuracy of the next frame will be better if the kernel size is increased; however, an increased kernel size raises the computational load. Our experiments show that a kernel size of 5×5 gives acceptable results balancing performance and computational load.

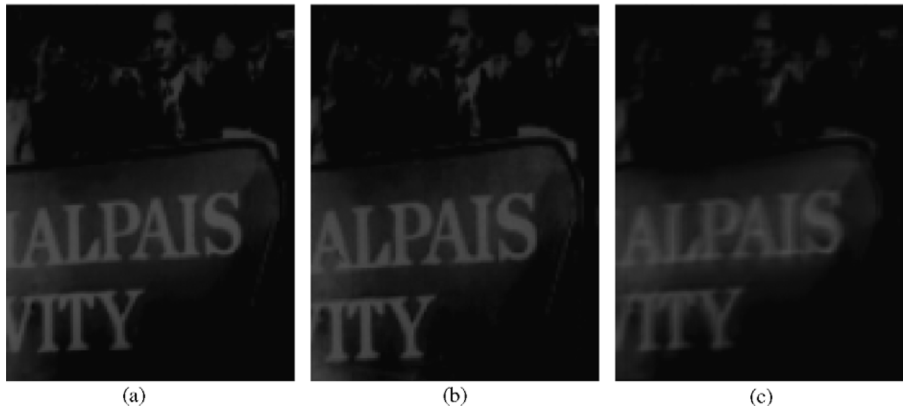

Fig. 2. (a) Mount #1028 (b) Mount #1029 (c) Estimation of (b) from (a) using best-fitting kernel.

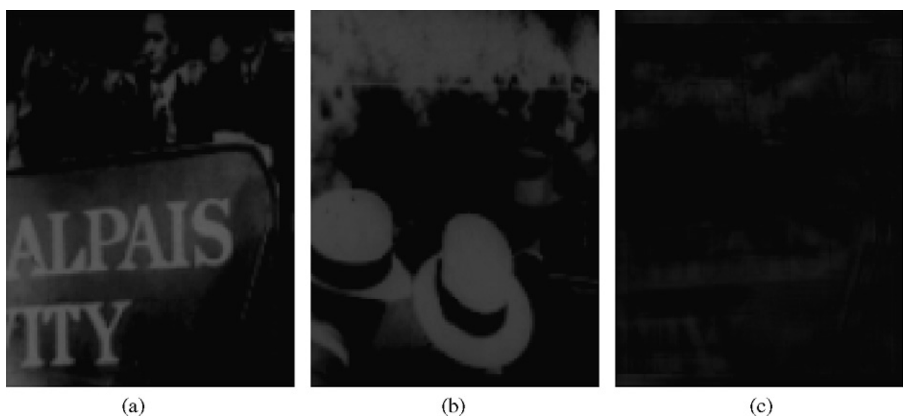

Fig. 3. (a) Mount #1029 (b) Mount #1030 (c) Estimation of (b) from (a) using best-fitting kernel.

The best-fitting kernel calculation is performed over sub-sampled image frames in order to further reduce the computational load. Sub-sampling by a factor 4 is suitable for this purpose. Sub-sampling operation is performed by simply calculating the mean value of neighbor pixels.

Two sub-sampled consecutive image frame from an archive film referred to as "Mount" (Mount Tamalpais Gravity Railroad – 1917) and the estimated second frame for similar and different scenes are given in Figs. 2 and 3, respectively. It can be seen from Figs. 2 and 3 that the estimation of the next frame from the previous frame using best-fitting kernel gives acceptable results for similar scenes. However, estimation of the next frame produces meaningless results in cases of scene-cuts.

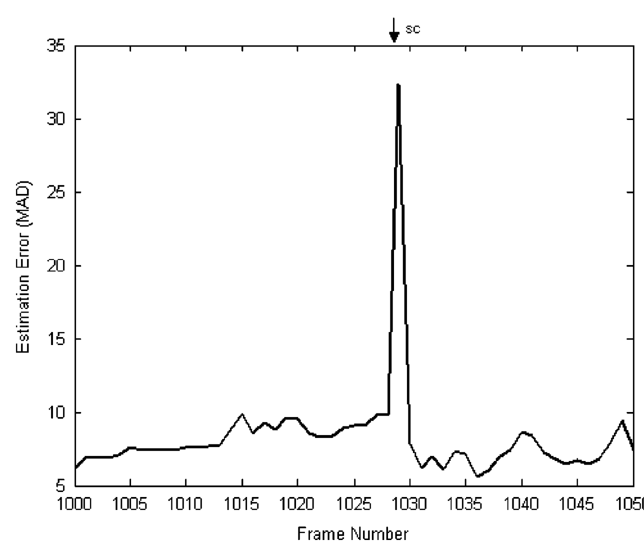

Fig. 4. Response of the decision metric in case of a scene-cut.

2.2. Decision metric and double thresholding approach

It is shown in the previous sub-section that the estimated frame is quite different for frames not belonging to the same scene. Success of the estimation can be measured using the absolute frame difference between estimated and original image frames as shown in (3)

$$\mathrm{DM}(t) = \frac{1}{w \times h} \sum_{i=0}^{w} \sum_{j=0}^{h} |\widehat{I}(i, j, t+1) - I(i, j, t+1)|. \quad (3)$$

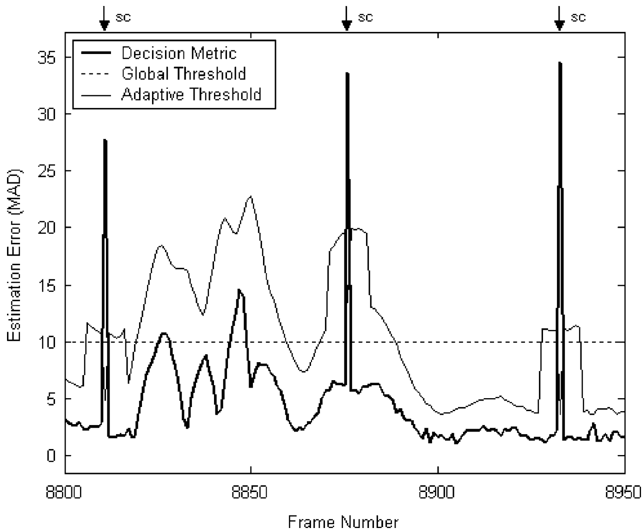

Fig. 5. Response of the decision metric in case of a scene-cut.

Response of the decision metric for a scene-cut of the "Mount" video is displayed in Fig. 4. The decision metric gives low values for similar scenes and has a peak in case of a scene-cut. Hence, a scene-cut decision can be given if the amplitude of the peak is higher than a certain threshold. The scene-cut (sc) instant is also depicted in the figure. The local adaptive threshold is computed as

if $\mathrm{DM}(t) < \mathrm{TH_G}$,

$$\mathrm{TH_{LL}}(t) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{DM}(t-i),$$

$$\mathrm{TH_{LR}}(t) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{DM}(t+i),$$

$$\mathrm{TH_L}(t) = \alpha\frac{\mathrm{TH_{LL}}(t)+\mathrm{TH_{LR}}(t)}{2}. \tag{4}$$

Here, $\mathrm{TH_G}$ denotes the global threshold, $\mathrm{TH_{LL}}$ denotes the left-hand site local threshold, $\mathrm{TH_{LR}}$ denotes the right-hand site local threshold, $\mathrm{TH_L}$ shows the final local threshold and N represents the length of the sliding window. Decision metrics of both sides of the candidate frame are evaluated and averaged using a scale factor (α) to obtain the local threshold value. If the decision metric value of the candidate frame is above the local threshold no scene-cut is decided, otherwise the frame is kept as candidate scene-cut and forwarded to the confirmation stage.

Fig. 5 shows the decision metric results of the proposed approach together with the global and local thresholds together with the scene-cut instants (sc) for a part of the "Birthoft" sequence. These results clearly demonstrate the advantage of the introduced double threshold strategy: while the estimation error values can be above the global threshold due to intensive local motion, the adaptive threshold avoids incorrect scene-cuts to be signaled and the three actual scene-cuts present in this part are detected successfully and passed to the confirmation stage.

2.3. Confirmation stage

One frame defects such as blotches and flicker (caused for instance from camera flashes) can decrease the estimation accuracy and cause false detections. A confirmation stage is therefore introduced into the proposed method to prevent these kind of errors. The estimation error between frame $(t-1)$ and frame $(t+1)$ or frame (t) and frame $(t+2)$ (one frame skip) will be low in the case of one frame defects. However, the estimation error will still be high if a scene-cut occurred. Thus, one frame defects are easily discarded from candidate scene-cuts evaluating the proposed measure with one frame skip for candidates without decreasing the correct scene-cut detection performance.

3. Experimental result

Three metrics used for the assessment of scene-cut detection algorithms are the recall, precision and $F1$ rates. The recall (R), precision (P) and $F1$ rates are defined as

$$R = \frac{C}{C+M}, \quad P = \frac{C}{C+F}, \quad F1 = \frac{2\times P\times R}{P+R}, \tag{5}$$

where C denotes the number of correctly detected scene-cuts, M denotes the number of missed scene-cuts, and F denotes the number of falsely detected scene-cuts. Table 1 compares the scene-cut detection performance of the proposed methods against techniques presented in the literature for a total of 10 archive films, and it is seen that the proposed approach provides a superior performance. These B&W archive films have various kinds of visual degradations as well as substantial camera and object motion. The archive films are obtained from free databases: The Open Video Project and Prelinger Archives.

A single-adaptive local threshold is employed for the histogram-based technique in [3] as well as the method proposed in [7], while the double threshold approach is utilized for the proposed method. The method proposed in [8] has its own automatic threshold. No specific threshold values is given in [6] so related thresholds are experimentally selected to give the best performance. The proposed best-fitting kernel-based approach outperforms all other techniques, and acceptable precision and recall rates are achieved for archive sequences that are in general difficult to segment due to visual degradations.

The method proposed in [8] appears to be better than our method in terms of recall. However, for our method the related threshold is chosen to give best overall performance

Table 1. Scene-cut detection performance comparison

Method Th. method	[3] Local	[6] Auto	[7] Local	[8] Auto	Prop. Double
C	804	855	815	975	939
M	205	154	194	34	70
F	339	161	103	200	67
P	70.34	84.15	88.78	82.98	93.34
R	79.68	84.74	80.77	96.63	93.06
$F1$	74.72	84.44	84.59	89.29	93.20

in terms of both precision and recall rates. The threshold parameters of the proposed methods can be tuned to optimize the recall rate. Therefore, the proposed method outperforms the other compared methods.

Missed hard-cuts are mainly caused by similar scene content. Intensive camera and objects motion result in false detections in some cases.

4. Conclusions

A novel scene-cut detection method based on the utilization of a best-fitting kernel between successive image frames is proposed for B&W archive films. The best-fitting kernel is employed to estimate the second frame of each successive image frame pair. The estimation error is calculated in terms of the absolute difference between original and estimated image frames. A double thresholding approach consisting of global and local adaptive thresholds is employed to decide candidate scene-cuts. A confirmation stage prevents false detections in case of single-frame defects commonly encountered in archive films. Superior performance of the proposed method is demonstrated with experimental results.

The proposed method mainly focuses on B&W archive films which have various visual degradations that can complicate the detection of hard-cuts and furthermore cause false detections. Methods proposed in the literature can fail in the case of B&W archive films and might not provide reasonable performance as shown in experimental results. Gradual transitions are not considered within this work, because most of the restoration methods can continue to work through scene transitions. However, hard-cut detection is crucial for restoration algorithms.

Acknowledgments

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant EEEAG/103E007. The authors would like to thank the Open Video Project (http://www.open-video.org) and Prelinger Archives (http://www.archive.org) for archive film material, as well as L.V. Gool and M. Osian for providing an executable version of the method of [8].

References

[1] Lefevre S, Holler J, Vincent N. A review of real-time segmentation of uncompressed video sequences for content-based search and retrieval. Real-Time Imag 2003;9: 73–98.
[2] Zhang H, Kankanhalli A, Simolar S. Automatic partitioning of full-motion video. Multimedia Systems 1993;1:10–28.
[3] Sethi I, Patel N. A statistical approach to scene change detection. In: SPIE conference proceedings on storage and retrieval for image and video databases III; 1995.
[4] Gargi U, Kasturi R, Strayer S. Performance characterization of video-shot-change detection methods. IEEE Trans Circuits Systems Video Technol 2000;10:1–13.
[5] Kim JR, Suh S, Sull S. Fast scene change detection for personal video recorder. IEEE Trans Consumer Electron 2003;49:683–8.
[6] Zabih R, Miller J, Mai K. A feature-based algorithm for detecting and classifying production effects. Multimedia Systems 1999;7:119–28.
[7] Vlachos T. Cut detection in video sequences using phase correlation. IEEE Signal Process Lett 2000;7:173–5.
[8] Osian M, Gool L. Video shot characterization. Mach Vision Appl 2004;15:172–7.

Oguzhan Urhan received his B.Sc., M.Sc., and Ph.D. degrees in Electronics and Telecommunications Engineering from the University of Kocaeli, Kocaeli, Turkey, in 2001, 2003, and 2006, respectively. His Ph.D. dissertation is about segmentation and restoration of archive videos. Since 2001, he has been with the Department of Electronics and Telecommunications Engineering, University of Kocaeli, as a Research Assistant. His research interests are in the area of digital signal and image processing, in particular image and video restoration and coding.

M. Kemal Gullu received his B.Sc., M.Sc., and Ph.D. degrees in Electronics and Telecommunications Engineering from the University of Kocaeli, Kocaeli, Turkey, in 2001, 2003, and 2006, respectively. His Ph.D. dissertation is about archive video restoration. Since 2002, he has been with the University of Kocaeli, Kocaeli, Turkey, where he is currently an Assistant Professor. His main research interests are in signal, image, and video processing, in particular image and video restoration and video stabilization.

Sarp Erturk received his B.Sc. degree in Electrical and Electronics Engineering from the Middle East Technical University, Ankara, Turkey, in 1995, and his M.Sc. degree in Telecommunication and Information Systems and Ph.D. degree in Electronic Systems Engineering from the University of Essex, UK, in 1996 and 1999, respectively. From 1999 to 2001, he carried out his compulsory service with the Army Academy, Ankara. Since 2001, he has been with the University of Kocaeli, Kocaeli, Turkey, where he is currently an Associate Professor. His research interests are in the area of digital signal and image processing.