IMPROVED PHASE CORRELATION BASED SCENE-CUT DETECTION FOR ARCHIVE FILM SEQUENCES

Oğuzhan Urhan | M.Kemal Güllü | Sarp Ertürk
e-mail: urhano@kou.edu.tr | *e-mail: kemalg@kou.edu.tr* | *e-mail: sertur@kou.edu.tr*
University of Kocaeli, Faculty of Engineering, Department of Electronics & Telecom. Engineering, 41040, Izmit, Kocaeli, Turkey

Key words: Cut detection, Phase Correlation, Video Indexing, Archive Film

ABSTRACT

This paper presents a novel robust scene-cut detection system based on modified phase-correlation. Initially, image frames are spatially sub-sampled. Next phase correlation based scene-cut detection is carried out for the sub-sampled image frames. A new double thresholding approach is originally proposed to detect candidate scene-cuts. Results confirm the superior performance.

I. INTRODUCTION

Scene-cut detection is usually the initial step and an important part of video segmentation, which has numerous applications in a variety of fields such as video retrieval, indexing, analysis, semantic description and compression. The development of shot boundary detection algorithms has been an important area of research as it is a necessity for nearly all video segmentation approaches and hence a prerequisite for higher level video content. Shot-boundary detection has also applications in other fields of video processing, such as video restoration for example. In the case of video restoration it is usually even more difficult to achieve a reasonable scene-cut detection performance as the video sequence usually contains visual degradations. This paper proposes a novel scene-cut detection technique that gives superb detection accuracy even for seriously distorted archive film sequences.

A video sequence is typically a collection of camera shots (or scenes) concatenated using postproduction techniques. An uninterrupted sequence of frames contiguously captured by the same camera is referred to as a single shot or scene. Transitions from one shot to the next can be classified into two basic types, namely abrupt scene-cuts and gradual transitions. Various methods have been proposed for the detection of scene-cuts and gradual scene changes [1-8]. Methods proposed for the detection of scene-cuts can be classified according to their main properties into pixel-based, histogram-based, feature-based and motion-based techniques [1]. A short literature review is given initially for the sake of completeness.

A pixel based method, which mainly considers pixel-wise differences of consecutive frames is proposed in [2]. The absolute sum of frame differences is used for detecting a scene-cut in this method. The grayscale histogram difference of successive frames using several metrics for scene-cut detection is considered in [3]. Color histograms in several color spaces such as RGB, HSV, YUV, Lab, and Luv have also been utilized 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. A feature based 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 which employs a two-dimensional affine model is utilized as a first step in [8]. The average pixel difference between motion compensated consecutive frames is thresholded using an adaptive threshold to detect scene-cuts in this work.

This paper proposes a novel scene-cut detection technique that is based on phase correlation, and the proposed technique is referred to as modified phase correlation based scene cut detection. The proposed technique introduces three important novelties: spatial image sub-sampling prior to phase correlation, double thresholding of phase-correlation peaks for scene-cut detection and false detection removal by mean and variance tests. The spatial sub-sampling is shown to improve robustness particularly against visual degradation and local motion. The double thresholding strategy consisting of a global threshold aided by an adaptive local threshold is shown to outperform the standard single threshold approach which uses either one. Finally a simple mean and variance test is utilized for heuristic exclusion of uniformly colored sequences in which cases phase correlation is extremely sensitive to noise. The main contribution of this paper is the whole solution, providing a robust and reliable scene-cut detection method. The proposed system is evaluated extensively and thoroughly and proven to give excellent scene-cut detection performance, even for visually degraded sequences, significantly outperforming previous approaches.

II. PROPOSED METHOD

In an ideal scene-cut, in which there is no similarity between the former scene and the new one, a high phase-correlation surface peak is obtained for frames belonging to the same scene; while a very low peak will result at the instant of the scene-cut where one frame belongs to the former and the next frame to the new scene. Therefore it is fundamentally possible to decide on the similarity of image frames according to the peak value in the phase-correlation surface and detect scene-cuts. In practical video sequences, however, scene-cuts are not always ideal. There might be similarities between two different shots because of common scenery, people or objects; or two different shots might have the same statistical features making it difficult to detect the scene-cut. Alternatively there might be substantial camera or object movement; noise or visual defects; or even brightness variations within a shot reducing similarities between consecutive frames of the same scene, potentially causing incorrect scene-cut decisions. While the phase-correlation based scene-cut detection proposed in [7] is stated to be more robust to brightness variations, noise and translational movements, it is demonstrated in the experimental evaluations that the technique is still likely to fail in the case of substantial camera or object movements; intensive flicker, rigorous noise and severe visual defects. In this paper, an enhanced phase-correlation based scene-cut detection technique outperforming previous approached is proposed.

PHASE CORRELATION OF SPATIALLY SUB-SAMPLED IMAGE FRAMES

If frame I_{k+1} is simply a spatially shifted version of I_k, then

$$I_{k+1}(x,y) = \alpha \times I_k(x-d_x, y-d_y) \tag{1}$$

where d_x and d_y show the horizontal and vertical displacements, and α represents a contrast difference. If $F_k(u,v)$ represents the two-dimensional discrete Fourier transform (DFT) of frame I_k, then the DFT of frame I_{k+1} will be

$$F_{k+1}(u,v) = \alpha \times F_k(u,v) e^{-j(ud_x+vd_y)} \tag{2}$$

In this case the phase-correlation surface is obtained as

$$S_{(k,k+1)} = F^{-1}\left[\frac{F_k(u,v)\times F_{k+1}^*(u,v)}{\left|F_k(u,v)\times F_{k+1}^*(u,v)\right|}\right] = F^{-1}\left[e^{j(ud_x+vd_y)}\right] \tag{3}$$

$$S_{(k,k+1)} = \delta(x+d_x, y+d_y)$$

Hence, the phase correlation surface will have a peak at the location corresponding to the displacement between the two images. A particularly useful feature of the phase correlation technique is the way performance degrades gracefully as conditions depart from the ideal of pure translation [9]. Identifiable peaks continue to be found provided that the global motion can be approximated by a translation; hence a small degree of global zoom and rotation, and even some amount of object movement within the frames can be compensated. However, if scene-cut detection based on phase-correlation is carried out and a scene-cut is decided according to the peak value of the phase correlation surface, it is possible that an incorrect scene-cut is signaled in the case of extensive zoom, rotation, noise or local motion.

The first contribution of this paper to phase-correlation based scene-cut detection is the process of spatial sub-sampling of image frames prior to the phase-correlation computation. It is proposed in this paper to spatially sub-sample image frames to reduce the effect of global zoom and rotation, local object motion, and visual degradations within image frames. Spatial sub-sampling of image frames is stated to have an inherent smoothing feature (see [10] for instance) and is therefore expected to reduce the effect of object and camera motions as well as visual degradations, rendering the image more suitable for detecting similarities between two images. Note that spatial sub-sampling is performed by averaging neighbor pixel values in the work presented in this paper to keep the computational cost of sub-sampling as low as possible, although the same performance is also observed for higher order approaches such as bicubic sub-sampling.

The effect of spatial sub-sampling on phase correlation is actually formulated in [11] to establish an extension of phase correlation to sub-pixel accuracy. It is shown in [11] that the phase correlation of sub-sampled image frames leads to a sub-sampled 2-D Dirichlet kernel which is very closely approximated by a 2-D sinc function. It follows that the effect of sub-sampling on the phase correlation process is a spread in the surface peak. However, an important aspect is that the main peak value does not change as it is independent of the sub-sampling factor (see eq. (18) in [11]), and therefore it can be argued that spatial sub-sampling of image frames is expected to have no effect on the main peak value (accordingly no negative effect), at least under ideal conditions.

Various tests are performed in order to observe that spatial sub-sampling of image frames has actually a positive effect on the main peak value, as the spatial sub-sampling process introduces robustness against noise and visual degradations, camera zoom and rotation, as well as local object motion. These experiments show that phase correlation peak values of the spatial sub-sampled images are always higher than that of original sized in the case of aforementioned situations.

Because spatial sub-sampling results in high frequency components being removed from the image frames, it is expected that images are made more equal and the phase correlation peak value should also increase for dissimilar images. Therefore it is also necessary to investigate the effect of spatial sub-sampling on dissimilar image frames (frames not belonging to the same scene) to ensure that a gain is achieved for scene-cut detection. For this purpose Fig. 1 displays the change in phase correlation peak value against sub-sampling factor for dissimilar images as well as image frames belonging to the same scene.

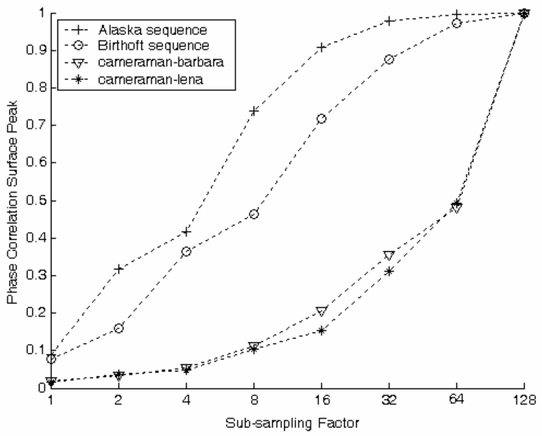

Figure 1. Effect of spatial sub-sampling on phase correlation.

It is seen that for a sub-sampling factor of two or four, an important increase in phase correlation peak value is obtained for similar frames (i.e. frames belonging to the same scene: successive frames of “Alaska” and “Birthoft” sequences). However, in the case of dissimilar frames (Cameraman-Lena and Cameraman-Barbara images) the increase in the phase correlation peak is rather limited. It is therefore easier to distinguish between similar and dissimilar frames after a spatial sub-sampling by a factor two or four. Naturally, if the image size is over-reduced (in the limit case image frames are sub-sampled to 1 pixel in each dimension) the phase correlation peak value increases significantly even for dissimilar frames, and in the limit case the phase correlation peak value reaches unity.

In summary, spatial sub-sampling of image frames by a small factor (of two or four), improves the ability to distinguish between similar and dissimilar frames using the phase correlation peak value.

DOUBLE THRESHOLD FOR SCENE-CUT DETECTION

The threshold strategy reported for scene-cut detection can be classified into two groups: global thresholds and local thresholds. Reported techniques typically use one of the two possible thresholding strategies. It has however been observed during the experimental work reported in Section III, that neither global thresholding nor local thresholding gives optimal performance, but it is possible to get an improved success if both are used together. Hence, it is therefore proposed in this paper to employ a double threshold strategy for phase correlation based scene-cut detection, consisting of a global threshold assisted by a local and hence adaptive threshold.

Our experiments show that a single global threshold might not give satisfactory detection results, yet a global threshold still provides a good measure for a candidate scene-cut. A local threshold on the other hand is useful to include local variations. It is therefore proposed in this paper to utilize a double threshold approach: a global threshold supported by an adaptive local threshold. In order to decide a scene-cut it is initially required that the phase correlation peak value falls below the global threshold. Secondly, a sliding window (of size $2w_s + 1$) local threshold is computed and a scene-cut decision is given if the phase correlation peak is also below this local threshold.

If a phase correlation peak value is obtained to be lower than the global threshold a sliding window is constructed comprising phase correlation peak values of w_s frames to the left-hand and w_s frames to the right-hand side of that particular candidate frame. Note that in the implemented approach, the window size is also defined to be adaptive, and the window is actually truncated at the first frame with a phase correlation peak value below a certain rate of the global threshold determined by a scale factor β . The mean phase correlation peak values of both sides of the candidate frame are computed and averaged and a scale factor (α) is used to obtain the local threshold value. If the phase correlation peak value of the candidate frames is above the local threshold no scene-cut is decided, otherwise the frame is kept as candidate scene-cut and forwarded to the false detection removal stage. The truncation of the sliding window at the point where the phase correlation value of the neighbor frames is below the global threshold ensures that no incorrect detection is accomplished if the phase correlation remains constantly very low for some time due to very intensive motion for instance. Note that in some cases it is possible that the phase correlation values of both adjacent frames are below the cut-off threshold ($\beta \times TH_G$) so that no sliding window can be constructed to compute the local threshold and in this case the local threshold is set to a very low value (typically 0.01) to avoid any incorrect detections.

if $p_{k,k+1} < TH_G$

$$TH_{LL}(k) = \frac{1}{N}\sum_{i=1}^{N} p_{k-i,k-i+1}, \{j\} = \left(j \middle| p_{k-j,k-j+1} < \beta TH_G, j = 1...w_s\right), N = \begin{cases} w_s & \text{, if } \{j\} = \varnothing \\ \min\ (j\text{-}1) & \text{, else} \end{cases}$$

$$TH_{LR}(k) = \frac{1}{N}\sum_{i=1}^{N} p_{k+i,k+i+1}, \{j\} = \left(j \middle| p_{k+j,k+j+1} < \beta TH_G, j = 1...w_s\right), N = \begin{cases} w_s & \text{, if } \{j\} = \varnothing \\ \min\ (j\text{-}1) & \text{, else} \end{cases} \quad (4)$$

$$TH_L(k) = \begin{cases} \alpha(TH_{LL}(k) + TH_{LR}(k))/2 & \text{if } TH_{LL} \neq 0 \text{ and } TH_{LR} \neq 0 \\ \alpha TH_{LL}(\text{k}) & \text{if } TH_{LL} \neq 0 \text{ and } TH_{LR} = 0 \\ \alpha TH_{LR}(\text{k}) & \text{if } TH_{LL} = 0 \text{ and } TH_{LR} \neq 0 \\ 0.01 & \text{if } TH_{LL} = 0 \text{ and } TH_{LR} = 0 \end{cases}$$

The computation of the adaptive local threshold can be summarized as given in (4). In this equation TH_G denotes the global threshold, TH_{LL} denotes the left-hand site local threshold, TH_{LR} denotes the right-hand site local threshold, TH_L shows the final local threshold computed, $p_{(k,k+1)}$ is the highest peak in the phase correlation surface for frames k and $k+1$, β is the scale factor determining how much below the global threshold value the window is to be truncated ($0 < \beta < 1$), and α is a scale factor specifying the local drop amount in the phase correlation peak value required for a scene-cut to be decided ($0 < \alpha < 1$).

Because an additional local threshold is employed, the global threshold can now be set to a value slightly higher than that would be used in a single global threshold strategy to reduce the number of misses. While the global threshold is slightly increased, in this case the local adaptive threshold ensures that the number of false detections is not raised. Optimum parameters for the local threshold computation process are determined experimentally, and the results presented in this paper are given for values of $TH_G = 0.08$, $ws = 5$, $\alpha = 0.25$, $\beta = 0.5$.

FALSE DETECTION REMOVAL BY MONITORING FRAME MEANS AND VARIANCES

Another modification to the phase-correlation based scene-cut detection process proposed in this paper is to monitor frame mean and variances to avoid incorrect detection resulting from noise and visual defects in single-colored image frames.

It has been observed that the phase-correlation based scene-cut detection incorrectly decides a scene-cut if noise or visual defects occur during single-colored image frames, which are for instance encountered during slow fade-in and fade-out effects. Due to the lack of spatial detail within the image, the phase-correlation method is highly sensitive to noise and visual defects in this case. Hence it is proposed to monitor frame variances and means, in parallel to the modified phase-correlation based scene-cut detection algorithm, and ignore scene-cuts signaled by the phase-correlation algorithm if frame variances remain below an extremely low threshold and the image mean shows only a small change, sustaining that the sequence continues to display the same single-colored image. In other words, an extremely low variance confirms that the image frame is single-colored (has very low spatial detail), and allowing only small changes in the intensity mean ensures that the same single-colored scene continues. Note that this step is executed only if a candidate scene-cut is signaled by the modified phase correlation process to confirm the scene-cut. Hence, a heuristic false detection removal procedure is being included into the phase correlation based scene cut detection process.

III. EXPERIMENTAL RESULTS

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

$$R = \frac{C}{C+M} \text{ and } P = \frac{C}{C+F} \quad (5)$$

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.

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 proposed in this paper is utilized for direct phase correlation based scene-cut detection (PC-SC) and the proposed MPC-SC. The method proposed in [8] has its own automatic threshold. The proposed MPC-SC technique 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 proposed MPC-SC misses only 4 scene-cuts and results in 14 false detections for a

Table 1. Scene-cut detection results for various test sequences with adaptive thresholds.

Method		[3]	PC-SC	[7]	[8]	MPC-SC
Threshold		Local	Double	Local	Auto	Double
TOTAL NUMBER OF SCENE-CUTS (1009)	C	804	855	815	975	1005
	M	205	154	194	200	4
	F	339	2211	103	34	14
	P	**70.34**	**27.89**	**88.78**	**82.98**	**98,63**
	R	**79.68**	**84.74**	**80.77**	**96.63**	**99,60**

total of 1009 scene-cuts encountered in 10 test sequences, and it is observed that the proposed double threshold approach reduces the number of missed scene-cuts as well as the number of false detections. The scene-cut detection method in [7] performs second in terms of precision and the method proposed in [8] performs third. Standard PC based scene-cut detection results in a very low precision rate with many false detections. In terms of the recall rate, the method proposed in [8] is second and PC based scene-cut detection is third, while the histogram based technique with χ^2 metric in [3] performs last.

IV. CONCLUSION

A novel scene-cut detection technique referred to as modified phase correlation based scene-cut detection has been proposed in this paper. The proposed Modified Phase Correlation based Scene-Cut (MPC-SC) detection method consists of three steps. Initially image frames are spatially sub-sampled and the peak phase-correlation value is computed. Then a double threshold strategy for scene-cut decision is accomplished. Finally, erroneous detection removal by variance and mean tests is carried out. Experimental results are demonstrated for archive film sequences that are particularly difficult to segment due to visual degradation, and it is shown that the proposed method provides superior detection performance. The spatial sub-sampling has a bonus benefit of substantially reducing the computational load required for the phase correlation process.

ACKNOWLEDGEMENT

This work was supported by the Turkish Scientific and Technical Research Council, TUBİTAK, under grand no. EEEAG103-007.

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. S. Lefevrea, J. Holler, and N. Vincent, A review of real-time segmentation of uncompressed video sequences for content-based search and retrieval, Real-Time Imaging, Vol. 9, pp. 73-98, 2003.
2. H. Zhang, A. Kankanhalli, and S. W. Smoliar, Automatic partitioning of full-motion video, Multimedia Systems, Vol. 1, pp. 10-28, 1993.
3. I.K. Sethi and N. Patel, A statistical approach to scene change detection, SPIE Conference Proceedings on Storage and Retrieval for Image and Video Databases III, San Diego/La Jolla, CA, USA, Vol. 2420, pp. 329-339, 1995.
4. U. Gargi, R. Kasturi, An evaluation of color histogram-based methods in video indexing, International Workshop on Image Databases and Multimedia Search, Amsterdam, The Netherlands, pp. 75–82, 1996.
5. J.R. Kim, S. Suh, S. Sull, Fast Scene Change Detection for Personal Video Recorder, IEEE Tran. on Consumer Electronics, Vol. 49, pp. 683-688, 2003.
6. R. Zabih, J. Miller, and K. Mai, A feature-based algorithm for detecting and classifying production effects, Multimedia Systems, Vol. 7, pp. 119-128, 1999.
7. T. Vlachos, Cut detection in video sequences using phase correlation, IEEE Signal Processing Letters, Vol. 7, pp. 173-175, 2000.
8. M. Osian and L.V. Gool, Video Shot Characterization, Machine Vision and Applications, Vol. 15, pp. 172-177, 2004.
9. S. Ertürk, and T. J. Dennis, Image sequence stabilisation based on DFT filtering,, IEE Proc. on Vision, Image, and Signal Processing, 147 (2), pp. 95-102, 2000.
10. B. L. Yeo, and B. Liu, Rapid scene analysis on compressed video, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 5, No:6, 1995, pp. 533-544, Dec 1995.
11. H. Foroosh, J. B. Zerubia, and M. Berthod, Extension of phase correlation to subpixel registration, IEEE Transaction on Image Processing, Vol. 11, pp. 188–200, 2002.