Modified Phase-Correlation Based Robust Hard-Cut Detection With Application to Archive Film

Oğuzhan Urhan, *Student Member, IEEE*, M. Kemal Güllü, *Student Member, IEEE*, and Sarp Ertürk, *Member, IEEE*

***Abstract*—This paper targets hard-cut detection for archive film, i.e., mainly black-and-white videos from the beginning of the last century, which is a particularly difficult task due to heavy visual degradations encountered in the sequences. A robust hard-cut detection system based on modified phase correlation is presented. Phase-correlation-based hard-cut detection is carried out using spatially sub-sampled video frames, and a candidate hard-cut is indicated in the case of low correlation. A double thresholding approach consisting of a global threshold used in conjunction with an adaptive local threshold is used to detect candidate hard-cuts. For uniformly colored video frames the phase correlation is extremely sensitive to noise and visual defects. Mean and variance based simple heuristic false removal at uniformly colored video frames is used at the final stage to prevent false detections in such cases. The paper provides a through theoretical analysis to show the usefulness of spatial sub-sampling. Furthermore through experimental results are presented for visual defects encountered in archive film material, to present the effectiveness of the proposed approach.**

***Index Terms*—Archive film, cut detection, phase correlation, video indexing.**

I. Introduction

HARD-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 analysis. 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 hard-cut detection performance as the video sequence commonly contains visual degradations. This paper presents a hard-cut detection technique that gives superior detection accuracy even for seriously distorted archive film sequences, such as typical black-and-white videos from the beginning of the last century.

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 hard-cuts (or hard-cuts) and gradual transitions. Various methods have been proposed for the detection of hard-cuts and gradual shot changes [1]–[37]. Methods proposed for the detection of hard-cuts can mainly be classified according to their basic properties into pixel-based, histogram-based, block-based, feature-based and motion-based techniques [1], [2]. A short literature review is given initially for the sake of completeness.

Pixel-based methods, which mainly consider pixel-wise differences of consecutive frames are proposed in [3], [4]. The global interframe difference of consecutive frames is used for detecting a hard-cut in [3]. It has been proposed to filter images to suppress camera motion and noise, and then count the number of significantly different pixels to decide a hard-cut in [4].

Histogram information of successive frames employing several metrics for hard-cut detection is considered in [5]. 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 [6]–[11]. While histogram-based techniques can provide a fairly reasonable performance for color sequences with minor visual defects and limited camera and object movements using color histogram information, the performance of histogram techniques is limited for grayscale sequences particularly in case of visually degraded frames.

Block-based evaluations have also been considered for pixel-based and histogram-based methods considering several color spaces [12]–[19]. A likelihood ratio, which is defined using block mean and variances is utilized to detect hard-cuts in [12]. Block mean differences in the HSV color space are computed using the hue and saturation components in [13]. Block sampled images in HSV color space are utilized in [14] to avoid illumination changes, and the deviation of block-based differences is used to detect hard-cuts. A local distance criteria defined in the RGB color space for blocks of consecutive video frames is employed in [15]. Block-based differences of successive frames in the RGB color space are utilized as main criteria to decide a hard-cut in [16]. Block-based histogram metrics in RGB, HSI and Lab color spaces are employed to decide hard-cuts in [17]–[19].

Several feature-based hard-cut detection methods have been proposed in [20]–[23]. Edge features of video frames are utilized for computing decision metrics in [20]–[23]. Motion compensation prior to edge detection is employed in [20] to gain robustness against motion. The method of [20] is improved in [21] by taking the edge information of several successive frames into account. The persistency of edge objects is utilized in [22]. Edge

Manuscript received May 13, 2005; revised October 6, 2005, and January 6, 2006. This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant EEEAG/103E007. This paper was recommended by Associate Editor R. Lienhart.

The authors are with the Department of Electronics and Telecommunications Engineering, University of Kocaeli, Kocaeli 41040, Turkey (e-mail: sarp@ieee.org).

Digital Object Identifier 10.1109/TCSVT.2006.875210

detection is performed in HSL color space to decide a hard-cut in [23]. It has been proposed in [24] and [25] to utilize a frequency domain correlation feature to detect hard-cuts. The technique proposed in [24] uses overlapping blocks and evaluates the phase correlation between each co-paired block in successive frames to decide a hard-cut in the case of low correlation. A similarity metric based on the frequency domain correlation of non-overlapping blocks is proposed to detect hard-cuts in [25].

Motion-based shot change detection methods make mainly use of the motion information to detect hard-cuts. The global motion information is acquired using parametric models in [26], and the consistency of global motion is utilized to detect a hard-cut. Multi-resolution motion estimation for global motion compensation which employs a two-dimensional affine model is utilized as a first step in [27] and [28]. Temporal evaluation of dominant motion is analyzed to decide shot changes in [27], whereas average pixel differences between motion compensated consecutive frames are thresholded using an adaptive threshold to detect hard-cuts in [28]. Motion compensation used in conjunction with a modified χ^2 test for histogram changes is proposed in [30]. Optical flow information is used to decide hard-cuts in [31] and [32]. Overlapping subsequences comprising three frames are used to estimate the next frame in [31], and a hard-cut decision is given if the actual frame is different from the prediction. In [32], video frames are initially sub-sampled and smoothened, followed by optical flow calculation and hard-cut detection based on likelihood computation for non-textured blocks.

The combination of several hard-cut detection techniques used together has also been utilized [33]–[37]. In [33] the χ^2 test for color histogram is utilized in conjunction with contour analysis to detect a hard-cut. The mean intensity value of the current frame, interframe pixel differences and changes in color distribution are employed with adaptive thresholds to detect a hard-cut in [34]. The method proposed in [35] uses mean intensity values, Euclidian distance, histogram comparison, likelihood ratio and motion estimation to decide a hard-cut. Thresholded pixel-wise and histogram differences are employed for hard-cut detection using a K-means clustering stage in [36]. DC image differences, χ^2 histogram test, and edge information are utilized together in a two phase approach to detect a hard-cut in [37].

Despite of the large amount of work carried out in this area, none of the previous techniques result in a robust and reliable approach that provides acceptable hard-cut detection performance particularly in the case of visually degraded sequences.

As stated in [24], phase-correlation-based hard-cut detection stands out in that it is fairly insensitive to the presence of global illumination changes and noise and is therefore stated to outperform established methods for cut detection in such cases. The robustness against illumination changes and noise is particularly important if video frames display visual degradation, which is commonly the case in archive film sequences for instance. Phase-correlation basically makes use of Fourier transform properties to detect similarities between video frames. The phase-correlation surface will have a high peak if two frames have high resemblance even in the case of a global translational shift (because the effect of translation is only an equivalent shift in the image phase), but a low peak will result if the similarity between consecutive frames is small, or the global motion cannot be resolved by simple translation only [38]. The phase-correlation-based cut-detection technique proposed in [24] uses overlapping blocks and evaluates the phase correlation between each co-paired block in successive frames and decides on a hard-cut in the case of low correlation. Although the approach of [24] is stated to perform better than histogram-based techniques due to its robustness against brightness variations and noise, experimental evaluations show that it is still affected by rigorous camera and object movements that cannot be approximated by trivial translational shifts, particularly camera zoom, as well as other visual defects that are commonly encountered in archive film.

This paper proposes a hard-cut detection approach referred to as modified phase-correlation (MPC) based hard-cut detection. The proposed approach consists of three main steps: phase correlation of spatially sub-sampled video frames, double thresholding of phase-correlation peaks for hard-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. Furthermore, a double thresholding strategy consisting of a global threshold aided by an adaptive local threshold is implemented to improve hard-cut detection performance. Finally, a simple mean and variance test is utilized for heuristic exclusion of false detections in uniformly colored frames in which cases phase correlation is extremely sensitive to noise. This paper provides a through theoretical analysis particularly for the effectiveness of spatial sub-sampling. The proposed approach is evaluated extensively and thoroughly, and is shown to give superior hard-cut detection performance, even for visually degraded sequences, outperforming previously reported techniques.

II. Modified Phase-Correlation-Based Hard-Cut Detection

If frame I_{k+1} is mainly 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

$$\begin{aligned} S_{(k,k+1)} &= F^{-1}\left[\frac{F_k(u,v) \times F^*_{k+1}(u,v)}{\left|F_k(u,v) \times F^*_{k+1}(u,v)\right|}\right] \\ &= F^{-1}\left[e^{j(ud_x + vd_y)}\right] \\ &= \delta(x + d_x, y + d_y) \end{aligned} \tag{3}$$

where $F^{-1}[]$ represents the inverse Fourier transform.

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 [38]. 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. In this case, the peak is usually reduced in amplitude and spatially extended. However, in the case of extensive zoom, rotation or local motion that cannot be approximated as a frame to frame translation it might not be possible to obtain easily identifiable peaks in the phase correlation surface. Hence, if hard-cut detection based on phase correlation is carried out and a hard-cut is decided according to the peak value of the phase correlation surface, it is possible that an incorrect decision is given in the case of extensive zoom, rotation, noise or local motion.

A high phase-correlation surface peak is typically obtained for frames belonging to the same shot due to similarities between consecutive frames. However, in an ideal hard-cut, where one frame belongs to the former and the next frame to the new shot, there is no similarity between the former shot and the new one and a very low phase correlation peak will be observed. Therefore, it is fundamentally possible to decide on the similarity of video frames according to the peak value in the phase-correlation surface and detect hard-cuts. In practical video sequences, however, hard-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 common content or similar statistical features making it difficult to detect the hard-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 shot, potentially causing incorrect hard-cut decisions.

It is important to note at this point that sometimes phase correlation is incorrectly being accounted to be equivalent to simple spatial domain correlation, with only the gain of reduced computational complexity. However, it is shown (see for instance [39]) that the peak in the phase correlation can be detected much more accurately compared to classical spatial domain cross correlation as phase correlation provides a distinct sharp peak in the case of correspondence, which is not the case in cross-correlation. Furthermore, the normalization inherent to phase correlation makes it more robust to noise correlated to the image functions, such as uniform illumination variations, average intensity offsets, blur, and fixed gain errors resulting from calibration. Hence, phase correlation will significantly outperform spatial domain cross correlation for hard-cut detection approaches.

A. Phase Correlation of Spatially Sub-Sampled Video Frames

The first stage of the approach presented in this paper consists of phase correlation of spatially sub-sampled video frames. Theoretical analysis is provided together with through experimental results to show that spatial sub-sampling is very effective against noise and visual degradations as well as camera and local motion. Spatial sub-sampling of video frames has an inherent smoothing feature (see [40] 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.

The effect of spatial sub-sampling on phase correlation is actually formulated in [39] to establish an extension of phase correlation to sub-pixel accuracy. It is shown in [39] that the phase correlation of sub-sampled video frames leads to a sub-sampled two-dimensional (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 [39, eq. (18)]), and therefore it can be argued that spatial sub-sampling of video frames is expected to have no effect on the main peak value (accordingly no negative effect), at least under ideal conditions.

In the following sub-sections it is demonstrated that spatial sub-sampling of video 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. An important bonus benefit of spatial sub-sampling is the reduction in the computational load, as reducing the size of video frames significantly decreases the computational load of the Fourier transform.

Anti-aliasing filtering before sub-sampling has been evaluated but it has been observed that the effect of anti-aliasing filtering on the phase correlation result is negligibly low and that hard-cut detection performance is not affected. Anti-aliasing filtering is therefore omitted to lower the computational load. Note that the negligibility of anti-aliasing filtering is also confirmed in the concluding remarks of [39].

The consequences of noise, camera rotation, camera zoom, or local object motion (i.e., non-matching parts) on the Fourier transform have already been well derived [39], [41]. Furthermore, the frequency domain counterpart of spatial sub-sampling is also well known. For phase correlation, the influence of these effects on the location of the phase correlation surface peak is also well derived. However, in the case of hard-cut detection, it is the phase correlation peak value amplitude that is being used to decide a hard-cut. In addition to establishing theoretical correspondences between the aforementioned effects and the phase correlation peak amplitude, various experimental results are presented in the following sub-sections to show the influence of spatial sub-sampling.

1) Influence of Noise: In order to investigate the influence of noise, it is assumed that frame I_{k+1} is affected by additive white Gaussian noise (AWGN), so that

$$I_{k+1}(x,y) = I_k(x,y) + n(x,y) \tag{4}$$

where $n(x,y)$ represents AWGN. Hence, the DFT of frame I_{k+1} is obtained as

$$\begin{aligned} F_{k+1}(u,v) &= F_k(u,v) + N(u,v) \\ &= G_k(u,v) + G_N(u,v) \\ &\quad + j[H_k(u,v) + H_N(u,v)] \end{aligned} \tag{5}$$

where $N(u,v)$ represents the 2-D DFT of AWGN, while $G(u,v)$ and $H(u,v)$ show the real and imaginary parts of the Fourier transform, respectively. In a similar approach to [42], if image pixels are assumed to be independent and identically

distributed (i.i.d) then the real and imaginary parts of the Fourier transform will be computed as

$$G_k(u,v) = \mathrm{Re}\left\{\sum_{x,y} I_k(x,y)e^{-j2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)}\right\} = \sum_{x,y} I_k(x,y)\cos\left[2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right] \quad (6)$$

and

$$H_k(u,v) = \mathrm{Im}\left\{\sum_{x,y} I_k(x,y)e^{-j2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)}\right\} = -\sum_{x,y} I_k(x,y)\sin\left[2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right]. \quad (7)$$

As a result of the central limit theorem, $G_k(u,v)$ and $H_k(u,v)$ will have Gaussian (normal) distributions. If $I_k(x,y)$ is zero mean (note that subtracting the mean value of the image will not affect the phase correlation process) and has a variance of $\mathrm{var}\{I_k(x,y)\} = E[I_k^2(x,y)] = \sigma_I^2$ (where $E[\cdot]$ represents expectation) then $G_k(u,v)$ and $H_k(u,v)$ will have zero means directly as a result of (6) and (7). Furthermore,

$$\begin{aligned}
&\mathrm{var}\{G_k(u,v)\}\\
&= \mathrm{var}\left\{\sum_{x,y} I_k(x,y)\cos\left[2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right]\right\}\\
&= E\left\{\sum_{x,y} I_k^2(x,y)\cos^2\left[2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right]\right\}\\
&= E\left\{\sum_{x,y} I_k^2(x,y)\frac{1}{2}\left(1+\cos\left[4\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right]\right)\right\}\\
&= \sum_{x,y} E\left\{I_k^2(x,y)\frac{1}{2}\left(1+\cos\left[4\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right]\right)\right\}\\
&= \sum_{x,y} E\left\{\frac{1}{2}I_k^2(x,y)\right\}\\
&\quad + \sum_{x,y} E\left\{\frac{1}{2}I_k^2(x,y)\cos\left[4\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right]\right\}\\
&= \frac{1}{2}\sum_{x,y} E\left\{I_k^2(x,y)\right\} + 0 = \frac{1}{2}MN\sigma_I^2.
\end{aligned} \quad (8)$$

A similar derivation can be performed for $H_k(u,v)$. Therefore, $G_k(u,v)$ and $H_k(u,v)$ will each have a variance equal to $(1/2)MN\sigma_I^2$.

If $n(x,y)$ has zero mean and a variance of $\mathrm{var}\{n(x,y)\} = \sigma_n^2$, then $G_N(u,v)$ and $H_N(u,v)$ will also be Gaussian (as a result of the central limit theorem because noise is by definition i.i.d.) with zero mean and variance $(1/2)MN\sigma_n^2$.

The phase-correlation surface is obtained as

$$\begin{aligned}
&S_{(k,k+1)}(x,y)\\
&= F^{-1}\left[\frac{F_k(u,v)\times F_{k+1}^*(u,v)}{\left|F_k(u,v)\times F_{k+1}^*(u,v)\right|}\right]\\
&= \frac{1}{MN}\sum_{u,v}\left[\frac{F_k(u,v)\times F_{k+1}^*(u,v)}{\left|F_k(u,v)\times F_{k+1}^*(u,v)\right|}\right]e^{j2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)}.
\end{aligned} \quad (9)$$

As there is no global displacement between the two frames in this case, the surface peak will be located at $(x,y) = (0,0)$ and the phase correlation peak value is obtained as

$$\begin{aligned}
&S_{(k,k+1)}(0,0)\\
&= \frac{1}{MN}\sum_{u,v}\left[\frac{F_k(u,v)F_{k+1}^*(u,v)}{|F_k(u,v)F_{k+1}^*(u,v)|}\right]\\
&= \frac{1}{MN}\sum_{u,v}\left[\frac{|F_k(u,v)|^2 + F_k(u,v)N^*(u,v)}{|F_k(u,v)||F_{k+1}(u,v)|}\right]\\
&= \frac{1}{MN}\sum_{u,v}\left[\frac{|F_k(u,v)|}{|F_{k+1}(u,v)|} + \frac{F_k(u,v)N^*(u,v)}{|F_k(u,v)||F_{k+1}(u,v)|}\right]\\
&= E\left[\frac{|F_k(u,v)|}{|F_{k+1}(u,v)|}\right] + E\left[\frac{F_k(u,v)N^*(u,v)}{|F_k(u,v)||F_{k+1}(u,v)|}\right].
\end{aligned} \quad (10)$$

Using the truncated Tylor series approximation, the expected value of the ratio of two random variables X and Y can be approximated as

$$E\left[\frac{X}{Y}\right] \approx \frac{\mu_X}{\mu_Y}\left[1+\left(\frac{\sigma_Y}{\mu_Y}\right)^2 - \frac{\sigma_{XY}}{\mu_X\mu_Y}\right] \quad (11)$$

where μ represents the mean, σ represents standard deviation and σ_{XY} represents the covariance between X and Y.

Because $F_k(u,v)$ and $N(u,v)$ are independent and the real and imaginary parts of $N(u,v)$ have zero mean values, the second term in (10) is obtained to be equal to zero, and therefore the phase correlation peak value is found to be

$$S_{(k,k+1)}(0,0) = E\left[\frac{|F_k(u,v)|}{|F_{k+1}(u,v)|}\right]. \quad (12)$$

The real part of $F_{k+1}(u,v)$ is obtained as $G_{k+1}(u,v) = G_k(u,v) + G_N(u,v)$. If $G_k(u,v)$ and $G_N(u,v)$ have Gaussian distributions, $G_{k+1}(u,v)$ will have a Gaussian distribution with zero mean and a variance equal to $(1/2)MN(\sigma_I^2+\sigma_n^2)$, which is also true for the imaginary part of $F_{k+1}(u,v)$.

Because the real and imaginary parts have Gaussian distributions with the same mean and variance, $|F_k(u,v)|$ and $|F_{k+1}(u,v)|$ will both have Rayleigh distributions. Therefore,

$$\begin{aligned}
E[|F_k(u,v)|] &= \sigma_I\sqrt{\frac{\pi}{2}}\sqrt{\frac{1}{2}MN} = \frac{\sigma_I}{2}\sqrt{\pi MN}\\
\mathrm{var}[|F_k(u,v)|] &= 0.429\frac{1}{2}MN\sigma_I^2 = 0.2145MN\sigma_I^2
\end{aligned} \quad (13)$$

and similarly

$$\begin{aligned}
E[|F_{k+1}(u,v)|] &= \frac{\sqrt{(\sigma_I^2+\sigma_n^2)}}{2}\sqrt{\pi MN}\\
\mathrm{var}[|F_{k+1}(u,v)|] &= 0.2145MN(\sigma_I^2+\sigma_n^2).
\end{aligned} \quad (14)$$

Hence, using (11) the phase correlation peak value is obtained as

$$\begin{aligned} &S_{(k,k+1)}(0,0) \\ &\quad\approx \frac{\mu_{|F_k(u,v)|}}{\mu_{|F_{k+1}(u,v)|}}\left[1+\left(\frac{\sigma_{|F_{k+1}(u,v)|}}{\mu_{|F_{k+1}(u,v)|}}\right)^2 \right. \\ &\quad\quad\left. -\frac{\sigma_{|F_k(u,v)||F_{k+1}(u,v)|}}{\mu_{|F_k(u,v)|}\mu_{|F_{k+1}(u,v)|}}\right] \\ &\quad\approx \frac{\sigma_I}{(\sigma_I+\sigma_n)}\left\{1+\frac{0.2145MN\left(\sigma_I^2+\sigma_n^2\right)}{\frac{\pi}{4}MN\left(\sigma_I^2+\sigma_n^2\right)}\right. \\ &\quad\quad\left. -\frac{\sigma_{|F_k(u,v)||F_{k+1}(u,v)|}}{\frac{\pi}{4}MN\sigma_I\sqrt{(\sigma_I^2+\sigma_n^2)}}\right\} \\ &\quad\approx \frac{\sigma_I}{(\sigma_I+\sigma_n)}\left\{1.273-\frac{4\sigma_{|F_k(u,v)||F_{k+1}(u,v)|}}{\pi MN\sigma_I\sqrt{(\sigma_I^2+\sigma_n^2)}}\right\}. \end{aligned} \tag{15}$$

The covariance $\sigma_{|F_k(u,v)||F_{k+1}(u,v)|}$ is defined as

$$\begin{aligned} &\sigma_{|F_k(u,v)||F_{k+1}(u,v)|} \\ &\quad= E[|F_k(u,v)||F_{k+1}(u,v)|] \\ &\quad\quad-\mu_{|F_k(u,v)|}\mu_{|F_{k+1}(u,v)|} \\ &\quad= E[|F_k(u,v)||F_k(u,v)+N(u,v)|] \\ &\quad\quad-\mu_{|F_k(u,v)|}\mu_{|F_{k+1}(u,v)|}. \end{aligned} \tag{16}$$

Because

$$\begin{aligned} &E[|F_k(u,v)||F_k(u,v)+N(u,v)|] \\ &\quad\le E[|F_k(u,v)|(|F_k(u,v)|+|N(u,v)|)] \end{aligned} \tag{17}$$

and

$$\begin{aligned} &E[|F_k(u,v)|(|F_k(u,v)|+|N(u,v)|)] \\ &\quad= E[|F_k(u,v)|^2+|F_k(u,v)||N(u,v)|] \\ &\quad= E[|F_k(u,v)|^2]+E[|F_k(u,v)||N(u,v)|] \\ &\quad= (E^2[|F_k(u,v)|]+\mathrm{var}[|F_k(u,v)|]) \\ &\quad\quad+E[|F_k(u,v)|]E[|N(u,v)|] \\ &\quad= \frac{\sigma_I^2}{4}\pi MN+0.2145MN\sigma_I^2 \\ &\quad\quad+\frac{\sigma_I}{2}\sqrt{\pi MN}\frac{\sigma_n}{2}\sqrt{\pi MN} \\ &\quad= \frac{\sigma_I^2}{4}\pi MN+0.2145MN\sigma_I^2 \\ &\quad\quad+\frac{\sigma_I\sigma_n}{4}\pi MN \end{aligned} \tag{18}$$

the worst case covariance can be obtained as

$$\begin{aligned} &\sigma_{|F_k(u,v)||F_{k+1}(u,v)|,\max} \\ &\quad= \frac{\sigma_I(\sigma_I+\sigma_n)}{4}\pi MN+0.2145MN\sigma_I^2 \\ &\quad\quad-\mu_{|F_k(u,v)|}\mu_{|F_{k+1}(u,v)|} \\ &\quad= \frac{\sigma_I(\sigma_I+\sigma_n)}{4}\pi MN+0.2145MN\sigma_I^2 \\ &\quad\quad-\frac{\sigma_I(\sigma_I+\sigma_n)}{4}\pi MN \\ &\quad= 0.2145MN\sigma_I^2 \end{aligned} \tag{19}$$

and in this case, the phase correlation peak value is obtained as

$$\begin{aligned} S_{(k,k+1)}(0,0) &\approx \frac{\sigma_I}{(\sigma_I+\sigma_n)} \\ &\quad\times\left\{1.273-\frac{4\left(0.2145MN\sigma_I^2\right)}{\pi MN\sigma_I\sqrt{(\sigma_I^2+\sigma_n^2)}}\right\} \\ &\approx \frac{\sigma_I}{(\sigma_I+\sigma_n)}\left\{1.273-\frac{0.273\sigma_I}{\sqrt{(\sigma_I^2+\sigma_n^2)}}\right\}. \end{aligned} \tag{20}$$

As $\sigma_I \ge 0$ and $\sigma_n \ge 0$, $(\sigma_I^2+\sigma_n^2) \le (\sigma_I^2+2\sigma_I\sigma_n+\sigma_n^2)$ with equality only valid if $\sigma_I = 0$ or $\sigma_n = 0$. Therefore, $\sqrt{(\sigma_I^2+\sigma_n^2)} \le (\sigma_I+\sigma_n)$ and hence $(\sigma_I/(\sigma_I+\sigma_n)) \le (\sigma_I/\sqrt{(\sigma_I^2+\sigma_n^2)})$. With an increase in the noise standard deviation, $(\sigma_I/(\sigma_I+\sigma_n))$ reduces obviously faster (takes a smaller value) compared to $(\sigma_I/\sqrt{(\sigma_I^2+\sigma_n^2)})$. Furthermore $(\sigma_I/\sqrt{(\sigma_I^2+\sigma_n^2)})$ is multiplied with a constant smaller than unity. Therefore, with an increase in the noise standard deviation, $(\sigma_I/(\sigma_I+\sigma_n))$ will decrease faster than $\{1.273-(0.273\sigma_I/\sqrt{(\sigma_I^2+\sigma_n^2)})\}$ increases. Hence, an increase in the noise standard deviation reduces the phase correlation peak value, while the peak value can be increased by reducing the noise standard deviation (Note that $S_{(k,k+1)}(0,0)=1$ for $\sigma_n=0$). In practice, pixel values are not i.i.d. but demonstrate a rather strong correlation, and therefore it is more appropriate to model the real and imaginary parts of the Fourier transform of an image using Generalized Gaussian Density functions just as it usually done for Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) coefficients and the magnitude of the Fourier transform can be modeled using the Weibull distribution which forms a generalization to the Rayleigh case [43]. However, if these models are used, the scale and shape parameters will vary specifically to image and noise characteristics as a result of which a general solution can not be obtained in a straightforward way. Hence, the simple Gaussian model is preferred to obtain a general idea of the relation between noise power and the phase correlation surface peak value.

A simple way to reduce noise power (noise standard deviation) is to perform signal averaging. When two AWGN components are added the resultant noise standard deviation increases by $\sqrt{2}$, while the signal power is doubled. Hence, the signal-to-noise ratio is expected to increase by $2/\sqrt{2}=\sqrt{2}$. For single video frames, it is possible to perform spatial averaging to reduce noise power. Because neighboring pixels will have some correlation, while noise is uncorrelated, averaging of neighboring pixels will increase the signal-to-noise ratio and hence the phase correlation peak value will increase with relation to (20).

In order to experimentally examine the influence of spatial sub-sampling on the phase correlation peak amplitude value under noise, synthetic distortions in the form of salt-and-pepper noise and Gaussian noise are introduced to test images. These effects are investigated because noise is typical to low quality video sequences (such as archive film material) and can affect the hard-cut detection performance, while these two types of noise reasonably express typical visual degradations encountered in archive film.

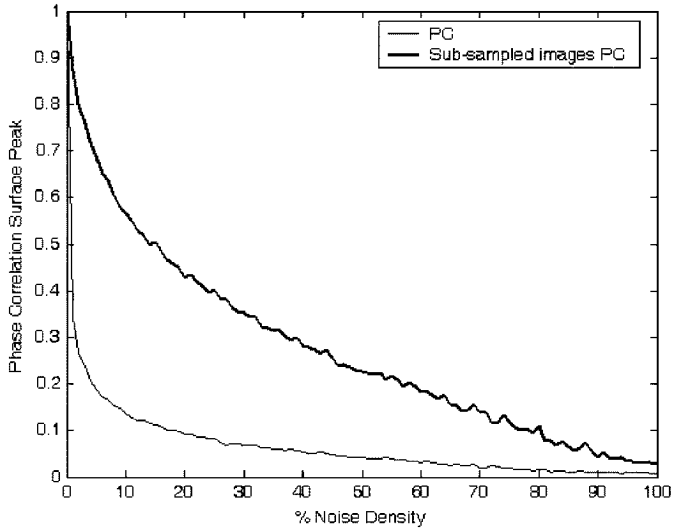

Fig. 1. Phase correlation surface peak values under salt-and-pepper noise.

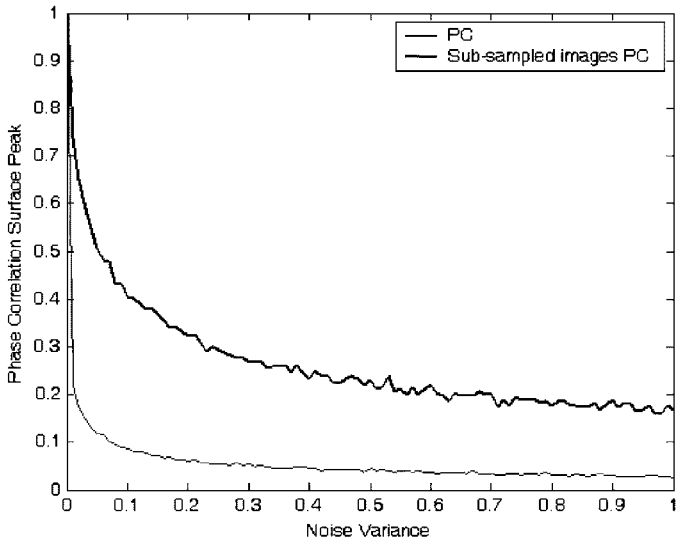

Fig. 2. Phase correlation surface peak values under Gaussian noise.

The phase correlation surface peak values obtained between the original Lena and salt-and-pepper noise added versions for various noise densities are displayed in Fig. 1. In order to investigate the influence of spatial sub-sampling, original and noisy images are spatially sub-sampled by a factor of four, and the phase correlation surface peak values obtained after spatial sub-sampling are investigated. Note that the horizontal axis of Fig. 1 indicates the noise density, while the vertical axis shows the phase correlation surface peak value. As observed from the figure, phase correlation surface peak amplitudes are increased about three times when using spatial sub-sampling in the case of salt-and-pepper noise. A particular important aspect is that the phase correlation peak value falls sharply after small noise densities if spatial sub-sampling is not employed, while the drop is much more gradual if spatial sub-sampling is utilized.

Phase correlation surface peaks obtained between the original Lena and Gaussian noise added versions as well as values obtained for sub-sampled Lena and sub-sampled noisy Lena images are plotted in Fig. 2. As observed from the figure, phase correlation surface peak amplitudes are increased about five times when using spatial sub-sampling in the case of Gaussian noise. Again the phase correlation peak value drops harshly after small noise powers if spatial sub-sampling is not employed, while the decrease is much more gradual in the case of spatial sub-sampling.

These results clearly demonstrate that spatial sub-sampling will add important robustness to the phase-correlation-based hard-cut detection process for noisy sequences, as the influence of noise is significantly reduced. This is an important advantage for sequences with intensive noise or visual degradations, as it is commonly the case for archive film sequences.

2) Influence of Flicker: Extreme lighting conditions and changes for instance occurring during a camera pointing to light source or flashes are likely to cause incorrect hard-cut decisions. If there is a contrast and brightness change between video frames, so that

$$I_{k+1}(x,y) = \alpha \times I_k(x,y) + \beta \tag{21}$$

then their Fourier transforms will be related by

$$F_{k+1}(u,v) = \alpha \times F_k(u,v) + \beta\delta(0,0). \tag{22}$$

In this case, the phase-correlation surface is obtained as

$$\begin{aligned} S_{(k,k+1)} &= F^{-1}\left[\frac{F_k(u,v) \times F^*_{k+1}(u,v)}{|F_k(u,v) \times F^*_{k+1}(u,v)|}\right] \\ &= F^{-1}\left[\frac{F_k(u,v) \times (\alpha \times F^*_k(u,v) + \beta\delta(0,0))}{|F_k(u,v) \times (\alpha \times F^*_k(u,v) + \beta\delta(0,0))|}\right] \end{aligned} \tag{23}$$

and therefore

$$\begin{aligned} S_{(k,k+1)} &= F^{-1}\left[\frac{\alpha \times |F_k(u,v)|^2 + \beta F_k(u,v)\delta(0,0)}{|\alpha \times |F_k(u,v)|^2 + \beta F_k(u,v)\delta(0,0)|}\right] \\ &= F^{-1}\left[\frac{\alpha \times |F_k(u,v)|^2 + \beta F_k(0,0)\delta(0,0)}{|\alpha \times |F_k(u,v)|^2 + \beta F_k(0,0)\delta(0,0)|}\right] \\ &= F^{-1}\left[\frac{\alpha \times |F_k(u,v)|^2 + \beta'\delta(0,0)}{\alpha \times |F_k(u,v)|^2 + \beta'\delta(0,0)}\right] \\ &= F^{-1}[1] \\ &= \delta(0,0) \end{aligned} \tag{24}$$

where $\beta' = \beta F_k(0,0)$ is used to change constants and $|\alpha \times |F_k(u,v)|^2 + \beta'\delta(0,0)| = \alpha \times |F_k(u,v)|^2 + \beta'\delta(0,0)$ because $|F_k(u,v)|^2$ is already real.

Although contrast and brightness changes are automatically cancelled in the phase correlation process under ideal situations, in practical cases there is usually a clipping effect. This results from luminance values exceeding the maximum allowed value (determined by the bit-depth) being clipped at the maximum.

In order to investigate the influence of spatial sub-sampling on the phase correlation result under brightness variations, synthetic global additive flicker is introduced to test images. Phase correlation surface peaks obtained between the original and global flicker (additive changes in luminance) introduced

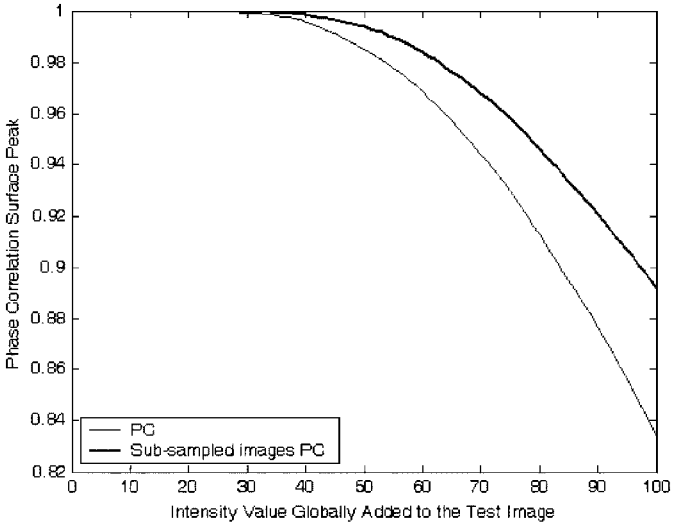

Fig. 3. Phase correlation surface peak values under brightness variations.

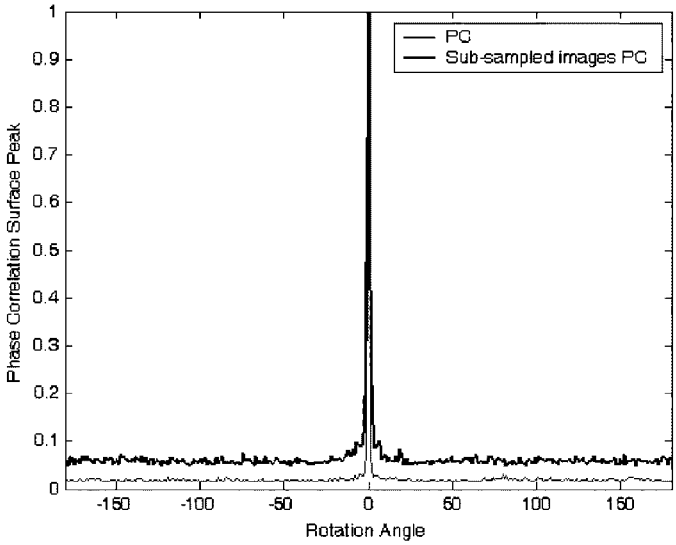

Fig. 4. Phase correlation surface peak values under camera rotation.

Lena images, as well as phase correlation surface peak values after sub-sampling are shown in Fig. 3. In this figure, the horizontal axis displays the intensity value globally added to the Lena image. It is observed from the figure that sub-sampling slightly increases phase correlation surface peak amplitudes particularly in the case of intensive flicker. The small gain can be explained as a result of pixel values being changed to a middle grey level because of the averaging process.

3) Influence of Camera Rotation and Zoom: If there is a rotation between video frames, so that

$$I_{k+1}(x,y) = I_k(x\cos\phi + y\sin\phi, -x\sin\phi + y\cos\phi) \quad (25)$$

then their Fourier transforms will show the same rotation and will be related by [44]

$$F_{k+1}(u,v) = F_k(u\cos\phi + v\sin\phi, -u\sin\phi + v\cos\phi). \quad (26)$$

In this case, the phase-correlation surface is obtained as

$$\begin{aligned} S_{(k,k+1)} &= F^{-1}\left[\frac{F_k(u,v)\times F_{k+1}^*(u,v)}{|F_k(u,v)\times F_{k+1}^*(u,v)|}\right] \\ &= F^{-1}\left[e^{j\theta_k(u,v)}e^{-j\theta_k(u\cos\phi+v\sin\phi,-u\sin\phi+v\cos\phi)}\right] \end{aligned} \quad (27)$$

where $\theta_k(u,v)$ represents the phase spectrum.

Because there is no displacement between the two frames in this case, the surface peak will be located at $(x,y) = (0,0)$ and the phase correlation peak value is obtained as

$$\begin{aligned} &S_{(k,k+1)}(0,0) \\ &\quad = \frac{1}{MN}\sum_{u,v}\left[\frac{F_k(u,v)F_{k+1}^*(u,v)}{|F_k(u,v)F_{k+1}^*(u,v)|}\right] \\ &\quad = E\left[e^{j\{\theta_k(u,v)-\theta_k(u\cos\phi+v\sin\phi,-u\sin\phi+v\cos\phi)\}}\right] \\ &\quad = E\left[\cos\{\theta_k(u,v) - \theta_k(u\cos\phi \right. \\ &\qquad \left. +v\sin\phi, -u\sin\phi + v\cos\phi)\}\right]. \end{aligned} \quad (28)$$

Note that the imaginary part can be ignored as the phase correlation result will be real. At small rotation angles, the effect of rotation is expected to be reduced by sub-sampling because pixel values will be averaged to intermediate values.

A similar analysis can be performed for scale changes. If frame I_{k+1} is a scaled version of I_k with scale factors (a,b) for the horizontal and vertical directions, their Fourier transform is related by [44]

$$F_{k+1}(u,v) = \frac{1}{|ab|}F_k(u/a, v/b). \quad (29)$$

In this case, the phase correlation peak value is obtained as

$$S_{(k,k+1)}(0,0) = E[\cos\{\theta_k(u,v) - \theta_k(u/a, v/b)\}]. \quad (30)$$

Particularly for small scale changes, the effect of scale change will be reduced by sub-sampling because pixel values will be averaged to intermediate values.

To investigate the effect of rotation, test images are initially rotated in counter clockwise direction with a step size of 1°. In order to avoid the effect of blank regions entering the image after rotation, 512×512 pixel sized images are rotated and the centre 256×256 pixel region is cropped afterwards. Phase correlation surface peak results against rotation angles for the Lena image are shown in Fig. 4, with and without spatial sub-sampling. It is noticed that phase correlation peaks of sub-sampled video frames are approximately three times higher, and sub-sampling is particularly useful for small rotation angles.

In order to examine the effect of zoom to the phase correlation process, the Lena image is zoomed from 1% to 100% with 1% steps using Bicubic interpolation. Phase correlation surface peak values against zoom rate are shown in Fig. 5. As observed from this figure, phase correlation peaks of spatially sub-sampled video frames are obtained to be approximately four times higher, and sub-sampling is again particularly useful for small scale changes.

These results clearly demonstrate that spatial sub-sampling will improve the robustness of the phase-correlation-based

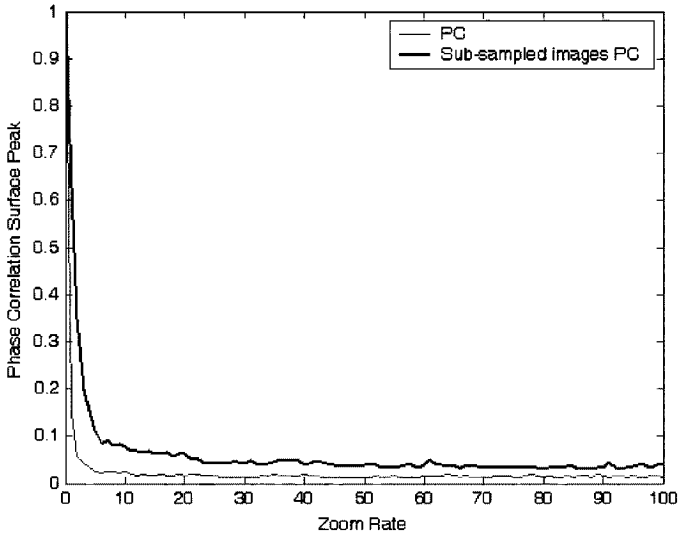

Fig. 5. Phase correlation surface peak values under camera zoom.

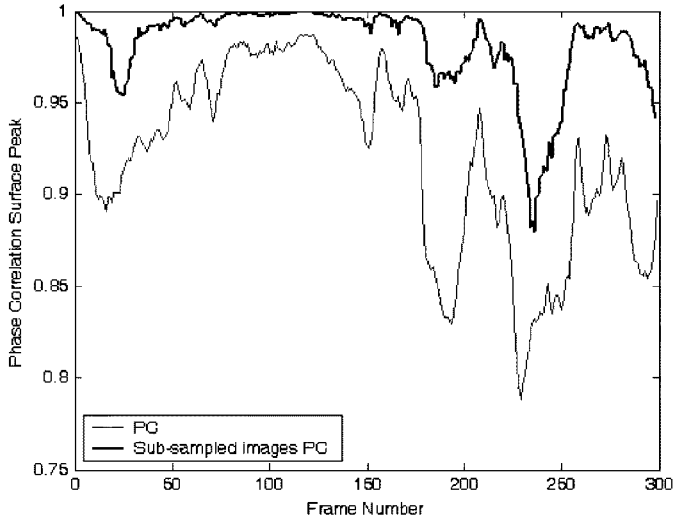

Fig. 6. Phase correlation surface peak values under local motion in the "Weather" sequence.

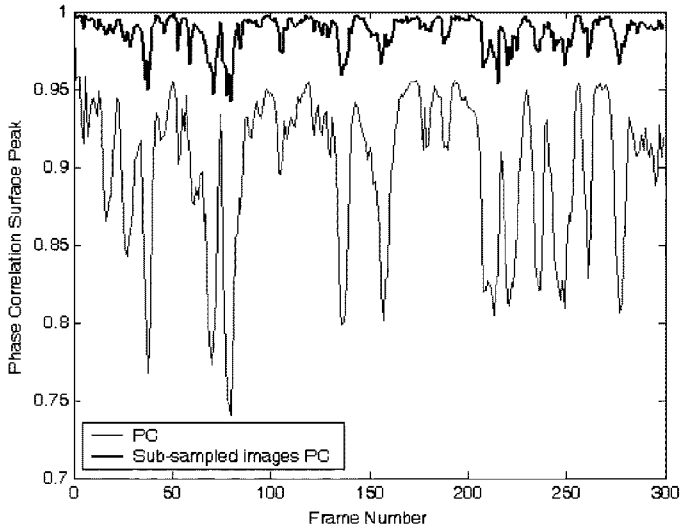

Fig. 7. Phase correlation surface peak values under local motion in the "Akiyo" sequence.

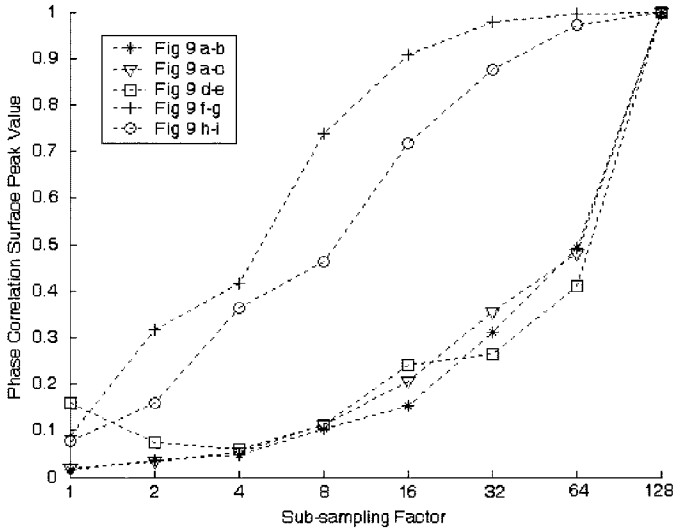

Fig. 8. Phase correlation peak values against sub-sampling factor for various similar and dissimilar images.

hard-cut detection process in the case of camera rotation and zoom. Although it is possible and has actually been proposed to utilize global motion compensation schemes within hard-cut detection systems to be more robust against rotation and zoom; motion estimation itself is likely to fail particularly for visually degraded sequences. It is therefore of advantage to utilize a technique that is not depending on motion compensation. The phase correlation approach inherently compensates for translational movements and as rotation or zoom is usually less effective, phase-correlation-based hard-cut detection provides a reasonable approach that does not require auxiliary global motion compensation. It is furthermore of advantage that spatial sub-sampling even improves the robustness against global camera motion particularly for small rotation and scale changes that are likely to occur in video sequences.

4) Influence of Local Motion: In case of local motion there will be non-overlapping parts or parts that display different displacements and hence result in secondary peaks reducing the height of the main peak value. Similar to [39], the total signal power can be defined as the squared ratio of the area of overlapping parts corresponding to the main peak, and the noise variance can be defined as the squared ratio of the area of non-overlapping parts. In this case, spatial sub-sampling will result in averaging of pixel values and is therefore expected to reduce the effect of non-overlapping parts.

In order to investigate the effect of spatial sub-sampling in the case of local motion, two well known test video sequences, namely the "Weather" and "Akiyo" test sequences are utilized. Both sequences are captured using a constant camera, and the "Weather" sequence contains negligibly low camera noise while the "Akiyo" sequence contains nearly no camera noise at all, so that the effect of local motion only can be assessed. As mentioned, these sequences are chosen as they basically do not contain any camera noise and camera motion, and are therefore very suitable for investigating the effect of local motion to the phase correlation process. Phase correlation surface peak values are shown in Figs. 6 and 7 for the "Weather" and "Akiyo" sequences, respectively. As observed from these figures, phase

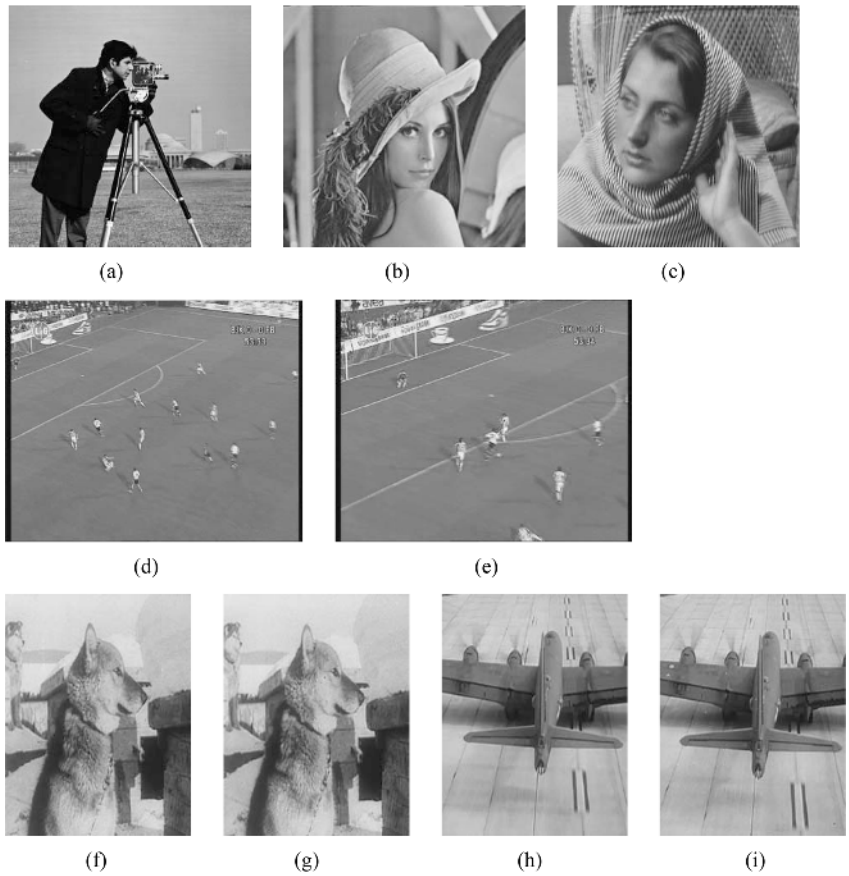

Fig. 9. Test images used in the sub-sampling factor results presented in Fig. 8. (a) Cameraman. (b) Lena. (c) Barbara. (d) Ball 1. (e) Ball 2. (f) Alaska # 05130. (g) Alaska # 05131. (h) Birthoft # 25255. (i) Birthoft # 25256.

correlation peaks of sub-sampled video frames are (sometimes significantly) higher, clearly illustrating that spatial sub-sampling introduces robustness against local motion.

5) Influence of Sub-Sampling on Dissimilar Images: Because spatial sub-sampling results in high frequency components being removed from the video 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 video frames (frames not belonging to the same shot) to ensure that a gain is achieved for hard-cut detection. For this purpose, Fig. 8 displays the change in phase correlation peak value against sub-sampling factor for dissimilar images as well as video frames belonging to the same shot. The images used for these results are given in Fig. 9. 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 shot). However, in the case of dissimilar frames 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 of two or four. It is furthermore seen from the results corresponding to the ball images, that spatial sub-sampling by a small factor can actually reduce the phase correlation peak value for two frames not belonging to the same shot but having similar content, which shows an important advantage for hard-cut detection. Naturally, if the image size is over-reduced (in the limit case video 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 video 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. Note that a sub-sampling factor of two or four has been found to be appropriate for videos with frame sizes of about MPEG1-2 sized frames (the archive films utilized in the results presented in this paper are of size 368×480 pixels while the MPEG-7 sequences have frame sizes of 320×240 pixels), but trials on high definition video (HDV) have shown that a higher sub-sampling factor (of eight or sixteen) can be utilized for larger frame sizes (videos with frame sizes of 1280×720 and 1920×1080 pixels are used for the corresponding HDV experiments).

B. Double Threshold for Hard-Cut Detection

Experimental work reported in Section III, shows that neither global thresholding nor local thresholding gives optimal performance, but it is possible to get an improved success if both are used together. Although a single global threshold might not give satisfactory detection results in the overall, it still provides a good measure for a candidate hard-cut in general. On

the other hand a local threshold is useful to include local variations. Hence, a double threshold strategy, consisting of a global threshold assisted by a local and hence adaptive threshold is utilized for phase-correlation-based hard-cut detection.

In order to decide a hard-cut it is initially required that the phase correlation peak value falls below the global 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 (i.e., the sliding window size is $2w_s + 1$). 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 the 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 frame is above the local threshold no hard-cut is decided, otherwise the frame is kept as candidate hard-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 a certain factor (β) 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 $(\beta \mathrm{TH}_G)$ 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.

The computation of the adaptive local threshold can be summarized as given in (31), shown at the bottom of the page. In this equation, TH_G denotes the global threshold, $\mathrm{TH}_{\mathrm{LL}}$ denotes the left-hand site local threshold, $\mathrm{TH}_{\mathrm{LR}}$ 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 hard-cut to be decided $(0 < \alpha < 1)$. Suitable parameters for the local threshold computation process are determined experimentally.

C. False Detection Removal by Monitoring Frame Means and Variances

An important problem with phase-correlation-based hard-cut detection is that the system incorrectly decides a hard-cut if noise or visual defects occur during single-colored video 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. It is again possible to use (20) to explain this effect. If the image standard deviation is very low (i.e., $\sigma_I \sim 0$) and the frame is affected by noise such that $\sigma_n \gg \sigma_I$ then it is possible to take $\sigma_I + \sigma_n \approx \sigma_n$ and $\sigma_I^2 + \sigma_n^2 \approx \sigma_n^2$ so that the phase correlation peak is obtained to be

$$S_{(k,k+1)}(0,0) \approx \frac{\sigma_I}{\sigma_n}\left\{1.273 - \frac{0.273\sigma_I}{\sigma_n}\right\} \approx 1.273\frac{\sigma_I}{\sigma_n} \approx 0. \tag{32}$$

Therefore, the phase-correlation-based hard-cut detection approach will signal an incorrect hard-cut in the case of noise if the frame variance is extremely low.

It is proposed to monitor frame means and variances, in parallel to the modified phase-correlation-based hard-cut detection algorithm, and ignore hard-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

$$\begin{aligned}
&\text{if } p_{k,k+1} < \mathrm{TH}_G \\
&\mathrm{TH}_{\mathrm{LL}}(k) = \frac{1}{N}\sum_{i=1}^{N} p_{k-i,k-i+1}, \\
&\{j\} = (j | p_{k-j,k-j+1} < \beta \mathrm{TH}_G, j = 1 \cdots w_s), \\
&N = \begin{cases} w_s, & \text{if } \{j\} = \emptyset \\ \min(j-1), & \text{else} \end{cases} \\
&\mathrm{TH}_{\mathrm{LR}}(k) = \frac{1}{N}\sum_{i=1}^{N} p_{k+i,k+i+1}, \\
&\{j\} = (j | p_{k+j,k+j+1} < \beta \mathrm{TH}_G, j = 1 \cdots w_s), \\
&N = \begin{cases} w_s, & \text{if } \{j\} = \emptyset \\ \min(j-1), & \text{else} \end{cases} \\
&\mathrm{TH}_L(k) = \begin{cases} \alpha(\mathrm{TH}_{\mathrm{LL}}(k) + \mathrm{TH}_{\mathrm{LR}}(k))/2 & \text{if } \mathrm{TH}_{\mathrm{LL}} \neq 0 \text{ and } \mathrm{TH}_{\mathrm{LR}} \neq 0 \\ \alpha \mathrm{TH}_{\mathrm{LL}}(\mathrm{k}) & \text{if } \mathrm{TH}_{\mathrm{LL}} \neq 0 \text{ and } \mathrm{TH}_{\mathrm{LR}} = 0 \\ \alpha \mathrm{TH}_{\mathrm{LR}}(\mathrm{k}) & \text{if } \mathrm{TH}_{\mathrm{LL}} = 0 \text{ and } \mathrm{TH}_{\mathrm{LR}} \neq 0 \\ 0.01 & \text{if } \mathrm{TH}_{\mathrm{LL}} = 0 \text{ and } \mathrm{TH}_{\mathrm{LR}} = 0 \end{cases}
\end{aligned} \tag{31}$$

TABLE I
DETAILS AND SOURCE OF ARCHIVE FILM SEQUENCES USED IN THE EXPERIMENTAL RESULTS

Short Name	Number of Hard Cuts	Full Archive Film Name and Year	Source
Alaska	74	Alaska's Silver Millions (Part I) (1936)	Prelinger Archives
Allinone	36	All in One (1938)	The Open Video Project
Allusa	68	All-American Soap Box Derby (1934)	The Open Video Project
Arteries	91	Arteries of New York City (1941)	The Open Video Project
Askme	206	Ask Me, Don't Tell Me (1961)	Prelinger Archives
Birthoft	302	Birth of the B-29 (1945)	The Open Video Project
Here	28	Here's Looking (1939)	Prelinger Archives
Mount	11	Mount Tamalpais Gravity Railroad (1917)	The Open Video Project
Panama	86	Panama-Pacific International Exposition (1940)	The Open Video Project
Social	107	Social Class in America (1957)	Prelinger Archives

single-colored image. In other words, an extremely low variance (below 50 for results presented in this paper) confirms that the video frame is single-colored (has very low spatial detail), and allowing only small changes in the intensity mean (below 20 for results presented in this paper) ensures that the same single-colored shot continues. Note that this step is executed only if a candidate hard-cut is signaled by the modified phase correlation process to confirm the hard-cut. Hence, a heuristic false detection removal procedure is being included into the phase-correlation-based hard-cut detection process.

III. EXPERIMENTAL RESULTS

The proposed hard-cut detection algorithm mainly relies on the phase correlation of spatially sub-sampled video frames. Spatial sub-sampling is performed by averaging neighbor pixel values of the original image (i.e., using the mean operation), which has an inherent smoothing effect. Fig. 10 shows the phase correlation peak values for original video frames as well mean and bicubic sub-sampled video frames (with a sub-sampling factor of four) for part of the "Birthoft" sequence, with hard-cuts being indicated. It is seen in Fig. 10 that the sub-sampling method (i.e., mean or bicubic) makes only a negligible difference, while sub-sampling clearly increases the phase-correlation peak value for frames belonging to the same shot.

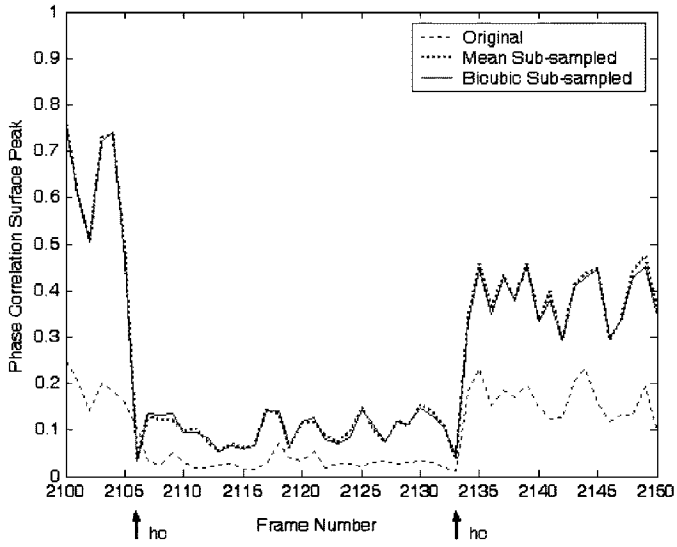

Fig. 10. Phase correlation peak values for original, mean sub-sampled, and bicubic sub-sampled frames of the "Birhoft" sequence between frames #2100–2150.

A. Experimental Setup

The performance of the proposed modified phase-correlation-based hard-cut detection algorithm, which consists of spatial sub-sampling of video frames and then double thresholding the phase correlation peak values followed by false detection removal using variance and mean monitoring is compared against several other hard-cut detection approaches. Two different sets of video sequences are used for the evaluation process. The first set consists of 10 archive film sequences, which are chosen to include all sorts of visual degradations encountered in archive film material as well as various effects such as large movement and abrupt illumination changes. Information about these sequences is provided in Table I. The second set consists of standard MPEG-7 test sequences, namely lgerca_lisa_1.mpg (54 hard-cuts) and lgerca_lisa_2.mpg (60 hard-cuts). These sequences are used as they contain numerous similar shots with related content and no distinctive visual degradation, and are captured by a hand held camera.

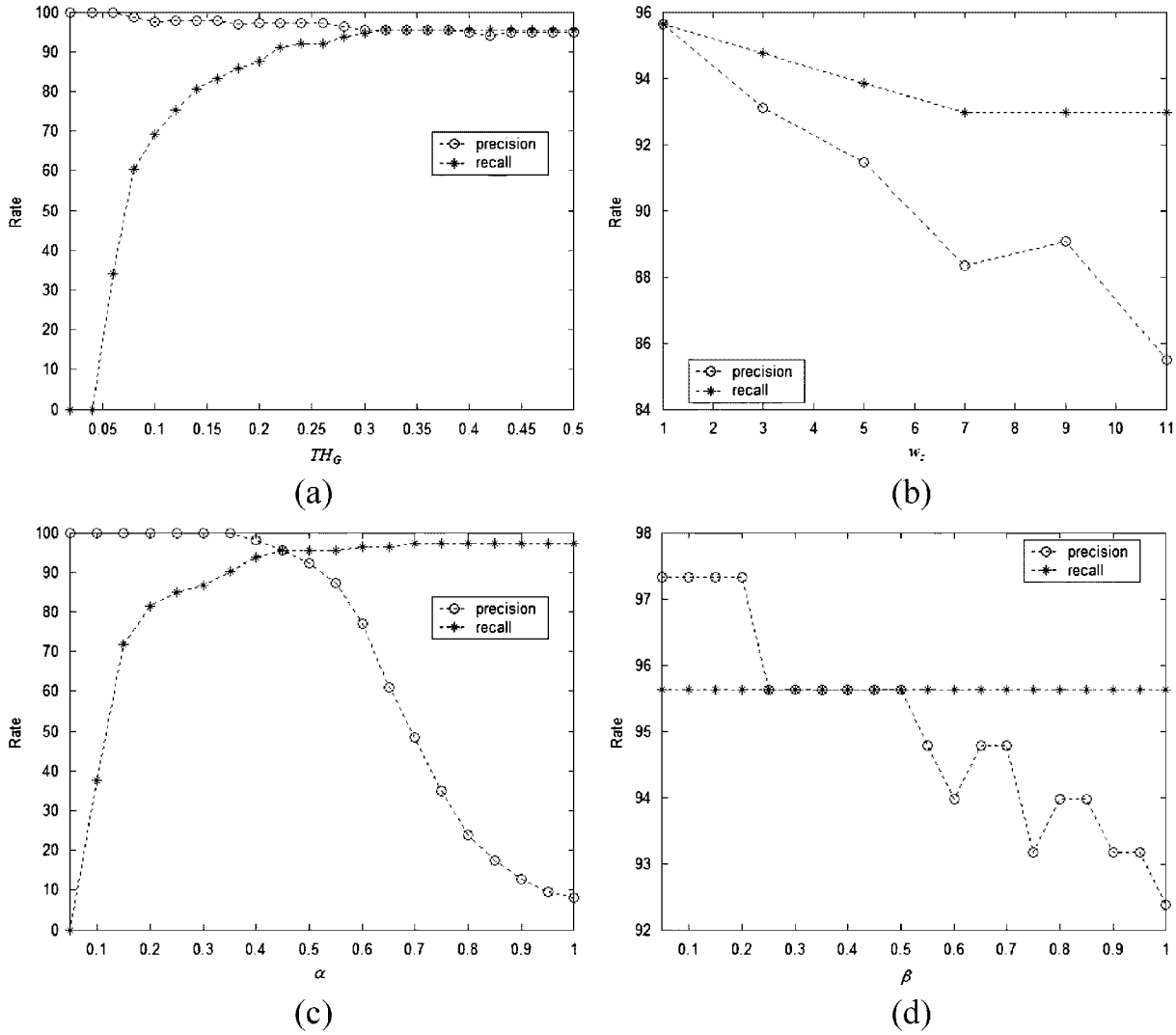

Fig. 11. Sensitivity of the proposed method with respect to introduced local threshold parameters. (a) Variable global threshold. (b) Variable windows size. (c) Variable α parameter. (d) Variable β parameter.

In order to assess the performance of hard-cut detection algorithms, two metrics, namely the recall rate and precision rate, are used. The recall rate (R) and precision rate (P) are defined as

$$R = \frac{C}{C+M} \quad \text{and} \quad P = \frac{C}{C+F} \tag{33}$$

where C denotes the amount of correctly detected hard-cuts, M denotes the amount of missed hard-cuts, and F denotes the amount of false detections. Recall and precision rates are typically inversely related; if detection parameters are changed to increase one the other typically decreases.

In order to determine the optimum threshold parameters of the double threshold strategy of MPC, all parameters of the double threshold (i.e., $\mathrm{TH}_G, w_s, \alpha, \beta$) are initially set to reasonable values and then each parameter is iteratively tuned, keeping the other three constant. Because recall and precision rates are inversely related, the optimum point is selected as the point where the recall rate is closest to the precision rate. The optimization is initially carried out using the archive film sequences only (i.e., set 1). In this case, the optimum parameters are obtained as $\mathrm{TH}_G = 0.15, w_s = 1, \alpha = 0.25$ and $\beta = 0.5$. While these parameters have been observed to also provide successful performance in terms of the precision rate for MPEG-7 sequences (i.e., set 2), the recall rate is relatively lower as some hard-cuts are missed due to similar shot content. For this reason the optimum parameters for the MPEG-7 set are also evaluated and obtained as $\mathrm{TH}_G = 0.32, w_s = 1, \alpha = 0.45$ and $\beta = 0.45$. In this case, the amount of missed hard-cuts is reduced in the MPEG-7 set, however, the amount of false detections increases. Fig. 11 shows the change in recall and precision rates for set 2, where always three of the parameters are kept at the optimum and the fourth parameter is varied to evaluate the sensitivity. It is seen from Fig. 11 that a small change in the parameters has only a slight effect on performance.

The first approach used for comparison is denoted as standard phase correlation (PC) that directly evaluates the phase correlation for entire video frames (i.e., without spatial sub-sampling) after which a hard-cut is signaled for low correlation values. A double threshold strategy is also utilized for PC-HC as it has been observed to give the best performance. The optimum double threshold parameters for PC-HC are obtained as $\mathrm{TH}_G = 0.02, w_s = 3, \alpha = 0.15$ and $\beta = 0.7$ for set 1, and

TABLE II
HARD-CUT DETECTION RESULTS WITH THRESHOLD PARAMETERS OPTIMIZED FOR ARCHIVE FILM SEQUENCES. CORRECT (C), MISSED (M), FALSE (F) HARD-CUTS, RECALL (R) AND PRECISION (P) RATES

Archive Films	Method	χ^2 [5]	HD [9]	Zabih [20]	PC-HC	OBBPC [24]	Osian [28]	MPC	MPC	MPC w/o HFR	MPC
	Threshold	Auto	Auto	Auto	Double	Double	Auto	Global	Adaptive	Double	Double
TOTAL (1009)	C	804	868	855	602	768	974	973	1002	1002	1002
	M	205	141	154	407	241	35	36	7	7	7
	F	369	235	161	407	229	200	33	15	84	13
	P	**68.54**	**78.69**	**84.15**	**59.66**	**77.03**	**82.96**	**96.72**	**98.53**	**92.27**	**98.72**
	R	**79.68**	**86.03**	**84.74**	**59.66**	**76.12**	**96.53**	**96.43**	**99.31**	**99.31**	**99.31**

MPEG-7 Sequences	Method	χ^2 [5]	HD [9]	Zabih [20]	PC-HC	OBBPC [24]	Osian [28]	MPC	MPC	MPC w/o HFR	MPC
	Threshold	Auto	Auto	Auto	Double	Double	Auto	Global	Adaptive	Double	Double
TOTAL (114)	C	88	85	99	40	91	110	20	97	97	95
	M	26	29	15	74	23	4	94	17	17	19
	F	37	24	84	0	22	21	0	0	0	0
	P	**70.40**	**77.98**	**54.10**	**100**	**80.53**	**89.97**	**100**	**100**	**100**	**100**
	R	**77.19**	**74.98**	**86.84**	**35.09**	**79.82**	**96.49**	**17.54**	**85.09**	**85.09**	**83.33**

Overall	Method	χ^2 [5]	HD [9]	Zabih [20]	PC-HC	OBBPC [24]	Osian [28]	MPC	MPC	MPC w/o HFR	MPC
	Threshold	Auto	Auto	Auto	Double	Double	Auto	Global	Adaptive	Double	Double
TOTAL (1123)	C	892	953	954	642	859	1084	993	1099	1099	1097
	M	231	170	169	481	264	39	130	24	24	26
	F	406	259	245	407	251	221	33	15	84	13
	P	**68.72**	**78.63**	**79.57**	**61.20**	**77.39**	**83.07**	**96.78**	**98.65**	**92.90**	**98.83**
	R	**79.43**	**84.86**	**84.95**	**57.17**	**76.49**	**96.53**	**88.42**	**97.86**	**97.86**	**97.68**

$\mathrm{TH}_G = 0.12, w_s = 11, \alpha = 0.35$ and $\beta = 0.05$ for set 2. Note that, the proposed false detection removal by variance and mean monitoring is also included in the PC-HC algorithms in order to reduce the number of incorrect detections so as to assess the best performance that can possibly be achieved by PC-HC without spatial sub-sampling but with double thresholding and heuristic false detection removal.

The second approach used for comparison makes use of the method proposed in [24] and is referred to as overlapped block-based phase correlation (OBBPC). In this case, video frames are divided into overlapping blocks of size 256×256 pixels with an overlap of 32 pixels for set 1 and an overlap of 16 pixels for set 2 (the second overlap is less as the frame sizes are smaller). Then the phase correlation peak for each co-sited block pair is computed and combined to give the detector response

$$R_{(k,k+1)} = 1 + \gamma \sum_{i=1}^{n} \log p^{(i)}_{(k,k+1)} \tag{34}$$

where n denotes the total number of blocks, $p^{(i)}_{(k,k+1)}$ is the highest peak in the phase correlation surface for the ith co-sited block pair, and γ is a scale factor used to normalize the detector response to provide an output in the range $(0, 1)$. Results are provided for a double threshold strategy with optimum parameters obtained as $\mathrm{TH}_G = 0.18, w_s = 3, \alpha = 0.5$ and $\beta = 0.3$ for set 1, and $\mathrm{TH}_G = 0.2, w_s = 3, \alpha = 0.5$ and $\beta = 0.95$ for set 2. Again, the proposed false detection removal by variance and mean monitoring is included in the OBBPC-HC algorithms to obtain the best possible performance.

Other than the two mentioned phase-correlation-based techniques, two commonly used histogram-based, namely χ^2 [5] and histogram difference (HD) [9] based hard-cut detection techniques are tested for comparison. The hard-cut detection performance of these techniques is evaluated for an adaptive thresholding approach, with the threshold value being computed as in [9], and the threshold parameter is obtained experimentally so as to give the best results. It has been observed that the false detection elimination proposed in this paper also works for histogram-based techniques and therefore the proposed false detection removal by variance and mean monitoring is also included into the histogram-based hard-cut detection algorithms to obtain the best possible performance.

Furthermore, the motion-based hard-cut detection technique proposed in [28] is also used for comparison to evaluate the performance of a technique that explicitly utilizes motion compensation.

Finally, the edge-based technique presented in [20] is evaluated, to assess the appropriateness of edge-based methods for archive film sequences. This approach has also been optimized for archive film sequences and MPEG-7 sequences, respectively.

B. Comparative Performance Evaluation

The recall and precision rates obtained for the test sequences of set 1 and set 2 are given in Tables II and III, respectively,

TABLE III
HARD-CUT DETECTION RESULTS WITH THRESHOLD PARAMETERS OPTIMIZED FOR MPEG-7 TEST SEQUENCES. CORRECT (C), MISSED (M), FALSE (F) HARD-CUTS, RECALL (R) AND PRECISION (P) RATES

Archive Films	Method	χ^2 [5]	HD [9]	Zabih [20]	PC-HC	OBBPC [24]	Osian [28]	MPC	MPC	MPC w/o HFR	MPC
	Threshold	Auto	Auto	Auto	Double	Double	Auto	Global	Adaptive	Double	Double
TOTAL (1009)	C	804	888	770	922	778	974	1007	1007	1007	1007
	M	205	141	239	87	231	35	2	2	2	2
	F	369	235	100	5228	245	200	1009	153	308	124
	P	**68.54**	**78.69**	**88.51**	**14.99**	**76.05**	**82.96**	**49.95**	**86.81**	**76.58**	**89.04**
	R	**79.68**	**86.03**	**76.31**	**91.38**	**77.11**	**96.53**	**99.80**	**99.80**	**99.80**	**99.80**

MPEG-7 Sequences	Method	χ^2 [5]	HD [9]	Zabih [20]	PC-HC	OBBPC [24]	Osian [28]	MPC	MPC	MPC w/o HFR	MPC
	Threshold	Auto	Auto	Auto	Double	Double	Auto	Global	Adaptive	Double	Double
TOTAL (114)	C	88	85	90	91	98	110	93	109	109	109
	M	26	29	24	23	16	4	21	5	5	5
	F	37	24	24	23	17	21	21	5	5	5
	P	**70.40**	**77.98**	**78.95**	**79.82**	**85.22**	**83.97**	**81.58**	**95.61**	**95.61**	**95.61**
	R	**77.19**	**74.98**	**78.95**	**79.82**	**85.96**	**96.49**	**81.58**	**95.61**	**95.61**	**95.61**

Overall	Method	χ^2 [5]	HD [9]	Zabih [20]	PC-HC	OBBPC [24]	Osian [28]	MPC	MPC	MPC w/o HFR	MPC
	Threshold	Auto	Auto	Auto	Double	Double	Auto	Global	Adaptive	Double	Double
TOTAL (1123)	C	892	953	860	1013	876	1084	1100	1116	1116	1116
	M	231	170	263	110	247	39	23	7	7	7
	F	406	259	124	5251	262	221	1030	158	313	129
	P	**68.72**	**78.63**	**87.40**	**16.17**	**76.98**	**83.07**	**51.64**	**87.60**	**78.10**	**89.64**
	R	**79.43**	**84.86**	**76.58**	**90.20**	**78.01**	**96.53**	**97.95**	**99.38**	**99.38**	**99.39**

for the proposed method as well as the other methods used for comparison.

Table II shows the hard-cut detection performance of the mentioned techniques with threshold parameters optimized for the archive film test sequences. It is observed that for archive sequences the proposed approach significantly outperforms all other techniques used for comparison. The proposed approach provides recall and precision rates of about 99%, whereas all other approaches provide significantly lower rates. It is seen that [9], [20], and [28] outperform the other phase-correlation-based approaches (PC-HC and [24]). If the hard-cut detection performance for MPEG-7 sequences is assessed with threshold parameters optimized for archive film, it is seen that the proposed approach provides an excellent precision rate (no incorrect hard-cut detections) but the recall rate drops. In this case, the recall rate of the proposed approach is lower than [20] and [28], and the main reason is that there are extremely similar shots in the second test set (shots in which filming is temporarily ceased and then the camera continues from nearly the same place) so that some hard-cuts are missed due to a low global threshold value. In the overall, the proposed approach outperforms all other methods used for comparison.

In order to evaluate the influence of each individual module of MPC-HC, results for MPC-HC are also provided for a single global threshold only, a single local threshold only, and without the heuristic false removal. Comparing PC-HC and MPC-HC with double thresholding it is seen that the additional spatial sub-sampling of MPC-HC provides a very important performance increase in terms of precision as well as recall. Comparing MPC-HC with double thresholding against single threshold approaches, it is seen that MPC-HC provides a good performance for any single threshold strategy, yet the double threshold performance is superior. While MPC-HC with only adaptive thresholding performs close to the double thresholding case, an additional benefit of the double threshold strategy is that a local threshold is only computed if the phase correlation peak value is below the global threshold, thus reducing the required computation amount. Comparing MPC-HC with the case where no heuristic false removal is utilized (MPC w/o HFR) it is seen that the introduced variance and mean tests successfully remove a significant number of false detections, hence the heuristic false removal strategy is very effective.

Table III shows the hard-cut detection performance of the mentioned techniques with threshold parameters optimized for the MPEG-7 test sequences. In this case, the proposed approach outperforms all other techniques for set 1, with only two real hard-cuts being missed due to the comparatively higher global threshold, however, the precision rate decreases. For the MPEG-7 sequences, the proposed approach misses only one more hard-cut than [28] but the precision rate is significantly higher as the number of incorrect detections is low. In the overall, the proposed approach again outperforms all other techniques used for comparison. Note that these results are important to show that the effectiveness of the proposed approach

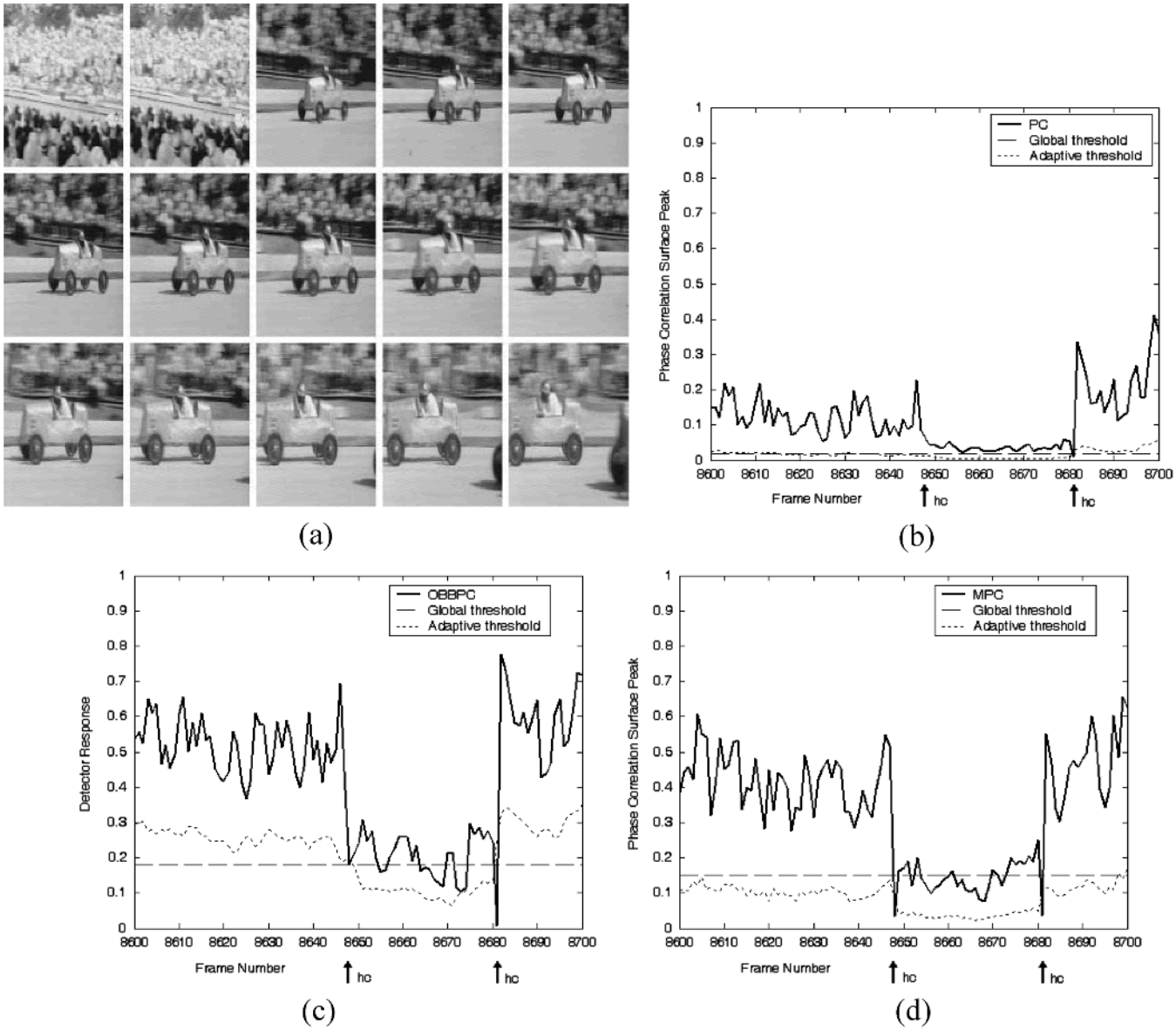

Fig. 12. (a) Frames # 8647–8661 of the "Allusa" sequence. (b) Phase correlation peak values and respective thresholds for part of the "Allusa" sequence for PC-HC. (c) Detector response values and respective threshold for part of the "Allusa" sequence for OBBPC-HC. (d) Phase correlation peak values and respective thresholds for part of the "Allusa" sequence for MPC-HC.

is not strictly depending on the optimization of threshold parameters. In this case, threshold parameters are optimized for sequences with totally different characteristics, however, the performance on archive films is still acceptable.

It has been observed that gradual shot change effects (such as dissolve, wipe, fade in and fade out) do not influence the overall hard-cut detection performance: none of the false detections occur during gradual transitions. The phase correlation process itself is robust to gradual transitions such as dissolves and wipes. While the frame variance drops to very low values during fade in and fade outs, in which case the phase correlation peak value can signal an incorrect hard-cut, as confirmed by (32), in this case, the heuristic false removal prevents any false detection. Hence, the proposed scheme suppresses gradual shot change effects successfully so that only hard-cuts are evaluated. The proposed approach is aimed at hard-cuts only and is obviously not suited to detect gradual shot changes, as the main aim is to suppress such effects to improve the hard-cut detection accuracy in the first place. For archive video restoration it is typically required to reset restoration algorithm parameters at the beginning of a new scene so that the restoration process will have a new start with no bias being introduced from previous scene characteristics. In the case of gradual changes it usually possible to continue running the restoration algorithm without any problem.

While theoretical analysis as well as precision and recall rates show that the presented approach is effective; two supplementary examples are presented in Figs. 12 and 13 for video parts of archive film sequences where the proposed MPC-HC technique outperforms the other phase-correlation-based PC-HC and OBBPC-HC methods. The figures also show the hard-cut (HC) instants, the adaptive local thresholds and the global thresholds

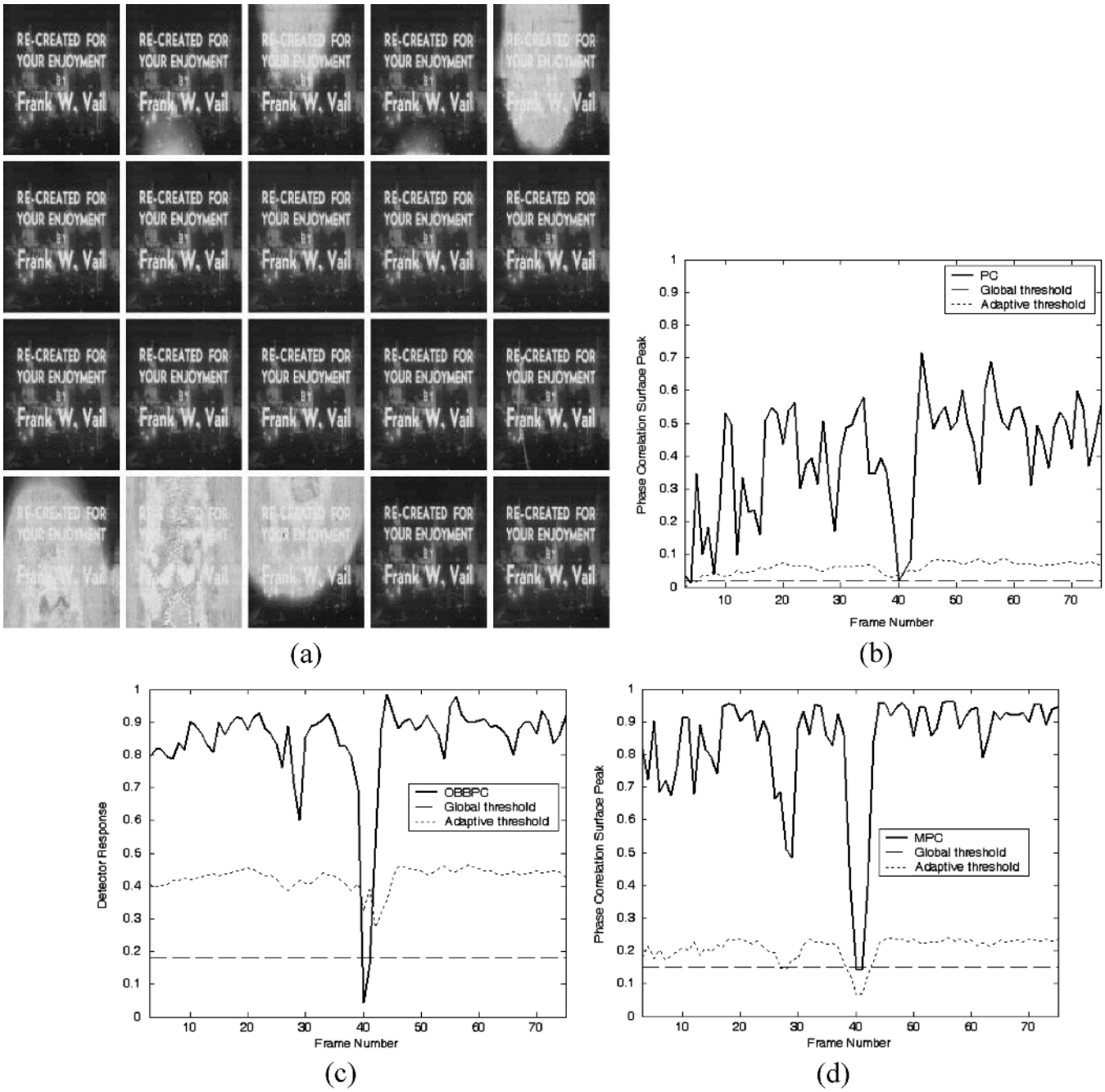

Fig. 13. (a) Frames # 27–44 of the "Panama" sequence. (b) Phase correlation peak values and respective thresholds for part of the "Panama" sequence for PC-HC. (c) Detector response values and respective threshold for part of the "Panama" sequence for OBBPC-HC. (d) Phase correlation peak values and respective thresholds for part of the "Panama" sequence for MPC-HC.

of the double threshold approach with parameters optimized for archive film performance. Fig. 12(a) shows part of the "Allusa" test sequence with intensive global camera motion and serious local object motion. As seen in Fig. 12(b), PC-HC misses one hard-cut and detects the other hard-cut correctly. Fig. 12(c) shows that this is also the case for OBBPC-HC, while furthermore one incorrect hard-cut is detected. However, MPC-HC detects both hard-cuts correctly as observed in Fig. 12(d). Fig. 13(a) shows part of the "Panama" test sequence illustrating serious visual defects (blotches) in various frames. While PC-HC and OBBPC-HC can signal these blotches as hard-cuts, in MPC-HC the phase correlation peak values are above the local threshold so that no incorrect decision is carried out in the proposed approach.

In summary, while the previously proposed OBBPC [24] improves the performance of hard-cut detection compared to standard PC, the performance is still insufficient and the computational load is significantly increased due to numerous overlapped FFT computations. The proposed modified phase-correlation-based hard-cut detection technique provides a superior detection performance with about 99% precision and recall rates even for difficult to segment archive films. Furthermore the

spatial sub-sampling considerably reduces the computational load of the FFT process. For a frame size of 368×480 pixels (a size typical to archive film sequences), for example, PC-HC requires the computation of a single 512×512 FFT (a power of two), OBBPC-HC requires the computation of 32 FFTs of size 256×256, while the proposed MPC-HC sub-samples the video frames to a size of 92×120 and therefore requires the computation of a single 128×128 FFT per video frame, and the same is valid for the IFFT computation encountered in the phase correlation process. The proposed hard-cut detection technique has therefore a considerably reduced computational load compared to the other phase-correlation-based detection algorithms and outperforms techniques previously proposed in literature such as the histogram-based techniques proposed in [5] and [9], the phase-correlation-based technique proposed in [24], the edge-based technique proposed in [20], as well as the motion-based technique proposed in [28].

IV. Conclusion

This paper presents a hard-cut detection approach referred to as modified phase-correlation-based hard-cut detection with application to b/w archive films. The presented approach consists of three steps. Initially video frames are spatially sub-sampled and the peak phase-correlation value is computed. Then a double threshold strategy for hard-cut decision is utilized. Finally, heuristic false detection removal by variance and mean tests is carried out.

The paper provides a through theoretical analysis to show the usefulness of spatial sub-sampling and heuristic false detection removal. Furthermore through experimental results are presented for visual defects encountered in archive film material, to present the effectiveness of the proposed approach. Results are demonstrated for archive film sequences that are particularly difficult to segment due to visual degradation, and it is shown that the proposed approach provides superior performance. The spatial sub-sampling has a bonus benefit of substantially reducing the computational load required for the phase-correlation-based detection process.

Acknowledgment

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 [28]. The authors would also like to thank the anonymous reviewers for their constructive feedback.

References

[1] S. Lefevre, 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, Feb. 2003.

[2] R. Lienhart, "Reliable transition detection in videos: A survey and practitioner's guide," *Int. J. Image and Graphics*, vol. 1, no. 3, pp. 469–486, Jul. 2001.

[3] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full-video search for object appearances," in *Proc. IFIP Working Conf. Visual Database Systems*, Budapest, Hungary, Oct. 1991, pp. 113–127.

[4] H. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video," *Multimedia Syst.*, vol. 1, pp. 10–28, Jul. 1993.

[5] I. K. Sethi and N. Patel, "A statistical approach to scene change detection," in *SPIE Conf. Proc. Storage and Retrieval for Image and Video Databases III*, La Jolla, CA, Feb. 1995, vol. 2420, pp. 329–339.

[6] D. Pye, N. J. Hollinghurst, T. J. Mills, and K. R. Wood, "Audio-visual segmentation for content-based retrieval," in *Proc. Int. Conf. Spoken Language Processing*, Sydney, Australia, Dec. 1998, pp. 1583–1586.

[7] B. Gunsel, A. M. Ferman, and A. M. Tekalp, "Temporal video segmentation using unsupervised clustering and semantic object tracking," *J. Electronic Imaging*, vol. 7, pp. 592–604, Jul. 1998.

[8] U. Gargi, R. Kasturi, and S. H. Strayer, "Performance characterization of video-shot-change detection methods," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 1–13, Feb. 2000.

[9] B. T. Truong, C. Dorai, and S. Venkatesh, "New enhancements to cut, fade, and dissolve detection processes in video segmentation," in *Proc. ACM Int. Conf. Multimedia*, 2000, pp. 219–227.

[10] W. Ren, M. Shamara, and S. Singh, "Automated video segmentation," presented at the 3rd Int. Conf. Information, Communications and Signal Processing (ICICS), Singapore, Oct. 2001.

[11] J. R. Kim, S. Suh, and S. Sull, "Fast scene change detection for personal video recorder," *IEEE Trans. Consum. Electron.*, vol. 49, no. 3, pp. 683–688, Aug. 2003.

[12] R. Kasturi and R. C. Jain, *Computer Vision: Principles*. Washington, DC: IEEE Computer Society Press, 1991.

[13] M. S. Lee, Y. M. Yang, and S. W. Lee, "Automatic video parsing using shot boundary detection and camera operation analysis," *Pattern Recogn.*, vol. 34, pp. 711–719, Mar. 2001.

[14] S. Lefevre, J. Holler, and N. Vincent, "Real time temporal segmentation of compressed and uncompressed dynamic colour image sequences," in *Proc. Int. Workshop on Real Time Image Sequence Analysis*, Oulu, Finland, Aug. 2000, pp. 56–62.

[15] C. H. Demarty and S. Beucher, "Morphological tools for indexing video documents," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, Florence, Italy, Jun. 1999, vol. 2, pp. 991–992.

[16] Y. F. Ma, J. Sheng, Y. Chen, and H. J. Zhang, "MSR-ASIA at TREC-10 video track: Shot boundary detection task," in *Proc. 10th Text Retrieval Conf. (TREC)*, Gaithersburg, MD, Nov. 2001, pp. 371–377.

[17] M. Ahmed and A. Karmouch, "Video segmentation using an opportunistic approach," in *Multimedia Modeling*, Ottawa, ON, Canada, Oct. 1999, pp. 389–405.

[18] M. Bertini, A. D. Bimbo, and P. Pala, "Content based indexing and retrieval of TV news," *Pattern Recogn. Lett.*, vol. 22, pp. 503–516, Apr. 2001.

[19] Y. Chahir and L. Chen, "Automatic video segmentation and indexing," in *Proc. SPIE Conf. Intelligent Robots and Computer Vision*, Boston, MA, Sep. 1999, vol. 3837, pp. 345–356.

[20] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying production effects," *Multimedia Syst.*, vol. 7, pp. 119–128, Mar. 1999.

[21] A. F. Smeaton, G. Gormley, J. Gilvarry, B. Tobin, S. Marlow, and N. Murphy, "An evaluation of alternative techniques for automatic detection of shot boundaries in digital video," in *Proc. Irish Machine Vision and Image Processing Conf.*, Dublin, Ireland, Sep. 1999, pp. 45–62.

[22] H. J. Heng and K. N. Ngan, "Integrated shot boundary detection using object-based technique," in *IEEE Int. Conf. Image Processing*, Kobe, Japan, Oct. 1999, vol. 3, pp. 289–293.

[23] T. Truong, S. Venkatesh, and C. Dorai, "Scene extraction in motion pictures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 1, pp. 5–15, Jan. 2003.

[24] T. Vlachos, "Cut detection in video sequences using phase correlation," *IEEE Signal Processing Lett.*, vol. 7, no. 7, pp. 173–175, Jul. 2000.

[25] S. Porter, M. Mirmehdi, and B. Thomas, "Temporal video segmentation and classification of edit effects," *Image Vis. Comput.*, vol. 21, pp. 1097–1106, Dec. 2003.

[26] M. Cherfaoui and C. Bertin, "Temporal segmentation of videos: A new approach," in *Proc. SPIE Conf. Digital Video Compression: Algorithms and Technologies*, San Jose, CA, Feb. 1995, vol. 2419, pp. 38–47.

[27] P. Bouthemy, M. Gelgon, and F. Ganansia, "A unified approach to shot change detection and camera motion characterization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 10, pp. 1030–1044, Oct. 1999.

[28] M. Osian and L. V. Gool, "Video shot characterization," *Machine Vis. Appl.*, vol. 15, pp. 172–177, Jul. 2004.

[29] B. Shahraray, "Scene change detection and content-based sampling of video sequences," in *Proc. SPIE Conf. Digital Video Compression: Algorithms and Technologies*, San Jose, CA, Feb. 1995, vol. 2419, pp. 2–13.

[30] T. Liu, X. Zhang, D. Wang, J. Feng, and K. T. Lo, "Inertia-based cut detection technique: A step to the integration of video coding and content-based retrieval," in *Proc. IEEE Int. Conf. Signal Processing*, Beijing, China, Aug. 2000, pp. 1018–1025.
[31] O. Fatemi, S. Zhang, and S. Panchanathan, "Optical flow based model for scene cut detection," in *Proc. Canadian Conf. Electrical and Computer Engineering*, Calgary, Canada, May 1996, vol. 1, pp. 470–473.
[32] W. K. Li and S. H. Lai, "Integrated video shot segmentation algorithm," in *Proc. SPIE Conf. Storage and Retrieval for Media Databases*, Santa Clara, CA, 2003, pp. 264–271.
[33] G. Quenot and P. Mulhem, "Two systems for temporal video segmentation," in *Proc. Eur. Workshop on Content Based Multimedia Indexing*, Toulouse, France, Oct. 1999, pp. 187–194.
[34] J. M. Gauch, S. Gauch, S. Bouix, and X. Zhu, "Real time video scene detection and classification," *Inf. Process. Manage.*, vol. 35, pp. 401–420, May 1999.
[35] Y. Yusoff, J. Kittler, and W. Christmas, "Combining multiple experts for classifying shot changes in video sequences," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, Florence, Italy, Jun. 1999, vol. 2, pp. 700–704.
[36] A. M. Ferman and A. M. Tekalp, "Efficient filtering and clustering for temporal video segmentation and visual summarization," *J. Visual Commun. Image Representation*, vol. 9, pp. 336–351, Dec. 1998.
[37] C. L. Huang and B. Y. Liao, "A robust scene-change detection method for video segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1281–1288, Dec. 2001.
[38] S. Ertürk and T. J. Dennis, "Image sequence stabilisation," *IEE Proc.-Vision, Image, and Signal Processing*, vol. 147, pp. 95–102, Apr. 2000.
[39] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–200, Mar. 2002.
[40] B. L. Yeo and B. Liu, "Rapid scene analysis on compressed video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 12, pp. 533–544, Dec. 1995.
[41] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2002.
[42] W. O. Saxton, "Transform noise statistics and Fourier component estimation," *Scanning Microscopy*, vol. 11, pp. 257–276, 1997.
[43] Q. Zhang and N. Boston, Statistical distribution index of modulation. [Online]. Available: http://www.math.wisc.edu/~boston/papers.html
[44] B. S. Reddy and R. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.

Oğuzhan Urhan (S'02) received the 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 Güllü (S'02) received the 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.

Currently, he is working as a Research Assistant in the Department of Electronics and Telecommunications Engineering, University of Kocaeli. His main research interests are in signal, image, and video processing, in particular image and video restoration and video stabilization.

Sarp Ertürk (M'99) received the B.Sc. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 1995, and the M.Sc. degree in telecommunication and information systems and the Ph.D. degree in electronic systems engineering from the University of Essex, U.K., 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.