

Automatic defect detection in video archives - Application to Montreux Jazz Festival Digital Archives

Philippe Hanhart^a, Martin Rerabek^a, Ivan Ivanov^a, Alain Dufaux^b, Caryl Jones^b,
Alexandre Delidais^b, and Touradj Ebrahimi^a

^aMultimedia Signal Processing Group, EPFL, Lausanne, Switzerland;

^bMetaMedia Center, EPFL, Lausanne, Switzerland

ABSTRACT

Archival of audio-visual databases has become an important discipline in multimedia. Various defects are typically present in such archives. Among those, one can mention recording related defects such as interference between audio and video signals, optical related artifacts, recording and play out artifacts such as horizontal lines, and dropouts, as well as those due to digitization such as diagonal lines. An automatic or semi-automatic detection to identify such defects is useful, especially for large databases. In this paper, we propose two automatic algorithms for detection of horizontal and diagonal lines, as well as dropouts that are among the most typical artifacts encountered. We then evaluate the performance of these algorithms by making use of ground truth scores obtained by human subjects.

Keywords: Video archives, defect detection, restoration, audio lines, dropouts

1. INTRODUCTION

The preservation of old audio and video recordings of major events that make history is essential for human knowledge. Over the last years, many efforts have been dedicated to digitalization and restoration of audio-visual archives, such as the Georges Méliès movie *A Trip to the Moon*, to preserve them from a certain death due to the aging of their storage formats. The first step consists in converting the old analog recording to a digital representation, which is better suited for nowadays storage, transmission, and processing technologies. This process is very delicate, as it requires special care to avoid damaging the physical support and, in the case of analog videotapes, to play the original content on a functional videotape player, which might be difficult to find and to properly maintain. Then, restoration can be applied on the digitized content, for example by removing defects due to the aging of the original physical support. The restoration can vary from a fully manual process, which is very expensive, to a fully automated process, which is very challenging as it requires detection and removal of defects without altering the original content. A few researchers have addressed this topic for the restoration of films and video archives.¹⁻⁴ In this paper, we present the results of our efforts on automatic defect detection in the framework of the restoration of the Montreux Jazz Festival Digital Archives.

The Montreux Jazz Festival was created by Claude Nobs in 1967. Thanks to the warm welcome offered to the artists, this festival immediately became a special one, a place where musicians could enjoy the venue in a relax and friendly atmosphere that gave rise to the famous Montreux jam sessions, those magical artistic moments when several musicians got together spontaneously, re-inventing their music in live duos or trios, on stage or somewhere in town. With in mind the vision of keeping a memory of the musical performances, and having a passion for technology and innovation, Claude Nobs decided to record the concerts from the first year of the festival. No less than 4'000 artists among the most talented musicians in the world came to Montreux, resulting in a unique live music collection of audio-visual recordings:

- 5'000 hours of video recordings, of which 2'000 are in critical condition
- 5'000 hours of audio recordings
- 4'000 documented concerts (metadata)
- 40'000 photos
- Hundreds of newspaper articles

Further author information: (Send correspondence to Philippe Hanhart) E-mail: {firstname.lastname}@epfl.ch

In Summer 2013, this archive was registered as part of the UNESCO program “Memories of the World”. Quincy Jones describes it as: “It is the most important testimonial to the history of music, covering Jazz, Blues, and Rock”. The entire original collection is preserved in a specially designed building with temperature and humidity control, and will be filed and preserved as part of the future Claude Nobs foundation.

In 2010, in partnership with Montreux Sounds SA, EPFL launched the Montreux Jazz Festival Digital Archives Project. The main motivation of the project is the preservation of this musical testimony, whose original storage media are subject to the destructive tests of time. To ensure access to it for future generations, an operation to digitize this unique catalogue is in process at EPFL, in collaboration with Vectracom in Paris (for digitization) and Amplidata in Belgium (for storage). This new reference archive and the tools setup around it at the MetaMedia Center (central information database, storage on hard drive, streaming, etc.) should allow fast and easy access to the data, enabling valorization and safeguard of the recordings for the future. Preservation and archive maintenance over the long term is the ultimate goal of the project. At the end of the Montreux Jazz Festival Digital Archives Project, the creation of a dedicated public space will allow visitors to experience a unique, personalized and immersive space to interact with the content of the archives. It will be a place where Art, Technology, and Culture combine. This space will open in 2015 and will be located at the heart of EPFL campus.

For EPFL, the originality and richness of the content make this archive a leading educational resource. More than 5'000 hours of video and audio recordings (in mono, stereo, or multi-tracks versions) are being made available to the researchers. This database is an invaluable digital resource that can be used in cutting-edge research for a wide range of areas such as computing, signal processing, acoustics, augmented reality, multimedia and security, to mention a few. Many EPFL laboratories are involved in the Montreux Jazz Festival Digital Archives Project, bringing their competences in archive valorization (metadata generation), defect detection, quality enhancement, or media transformation for use in new display technologies. In particular, the Multimedia Signal Processing Group of EPFL is responsible for the automatic defect detection in the video archives. Based on a detailed analysis of the digitized content, a set of typical defects was established. Among them, the audio lines and dropouts were selected because of their frequency of occurrence and impact on perceived quality. Fully automatic detection algorithms were developed to detect these two defects. In this paper, we describe these algorithms in details and report their performance evaluated using a ground truth database of video sequences annotated by human subjects. There are several solutions proposed in the literature dealing with detection of various artifacts. Below, we focus on two most disturbing artifacts, namely quasi-periodic noise (audio lines) and line or pixel drop (dropouts), therefore, we present here the overview of current work related to them only.

The goal of works found in literature focusing on periodic and quasi-periodic noise is its removal. Until now, two different solutions have been introduced, namely Wiener filter and masking to eliminate related spectral peaks. The Wiener filter,⁵ albeit its computational complexity, performs well only in cases when a proper and precise noise model is known. A second approach exploits the advantages of Fourier amplitude spectrum properties to detect a periodic noise. Strictly periodic noise in spatial domain corresponds to sharp peaks in Fourier domain and can be easily localized using thresholding techniques.⁶ These simple methods can be used on a frame-by-frame basis where for each frame a new threshold value has to be set. Moreover, in practical cases, the noise is not purely periodic and affects all surrounding spectral coefficients around the peaks. Therefore, the use of such approaches is limited and not practical for processing large number of images. A spectral peak detector dealing with quasi-periodic noise has been introduced in⁷ where authors compare amplitude spectral components to local spectrum median to identify the peaks. This method shows very good performance, however, using median filter is relatively slow, especially for large window sizes, which is crucial for high resolution content (SD and higher). This, of course, assumes that the sampling of Fourier transform corresponds to the image resolution. Processing a huge amount of data, as it is the case of Montreux Jazz Festival Digital Archives, requires relatively fast automatic methods. Therefore, using median filter is not suitable. Below, we introduce a fast automatic technique that relies on mean and standard deviation as statistical properties of Fourier amplitude spectrum. This method, unlike the previous one, can exploit the summed area table,⁸ which assures high speed computations.

Dropouts are a common problem in video archives. Despite the importance of dropout defects, few publications directly deal with the detection of this artifact. Abbas and Domanski⁹ have proposed an approach that

uses a nonlinear filter (e.g., three-dimensional median filter with motion compensation) as a predictor of a pixel's corruptness and uses the prediction error to control the restored pixel. The restored pixel takes the value of either the input pixel itself, the nonlinear filter output, or a linear combination of both. In turn, the approach by Harvey and Marshall¹⁰ is based on rearranging of pixel data into a format similar to that in which the original analog data would have been found on the videotape. This technique brings the scratched data into alignment, where corrupted pixels can be accurately detected by making use of morphological operators and thresholding. An approach for severe dropouts detection has been proposed by Kaprykowsky *et al.*¹¹ This technique uses a two-stage classification based on global color statistics of the frames for defect detection. Similar works deal with detection of vertical line scratches in video sequences.^{12,13} These scratches mostly remain in the neighborhood of the same location over frames, unlike dropouts, which are randomly spread over a frame across the time, thus making them more challenging for detection. However, most of the prior work on dropouts processing skips the detection part, assuming that frames with such an artifact are already known (e.g., manually selected), and mostly perform some form of global filtering for dropouts removal.

The remainder of this paper is organized as follows. A brief overview of the digitalization process of the Montreux Jazz Festival Digital Archives and a description of the typical defects encountered in the digitized content are given in Sec. 2. The algorithms developed to automatically detect the audio lines and dropouts are described in Sec. 3. In Sec. 4, the performance of the algorithms is reported and analyzed. Finally, concluding remarks and discussion on future work are given in Sec. 5.

2. VIDEO ARCHIVES

2.1 Digitalization

Many different tape formats were used across the years of the festival. The festival co-produced with the Télévision Suisse Romande its first television programs. At the beginning, analog 2-inch tapes were a major format. Then, spotting the latest audio-visual technology at the TV Symposium of Montreux, the festival quickly switched from black-and-white to color (1970), from stereo to multi-tracks (1975), and from standard definition to high definition (1991). The result is a vast catalog of concerts recorded and stored on different formats. For many years, several formats were used in parallel, exploiting conventional technology of the time, as well as emerging technologies on the rise.

One of the main challenges was the speed of digitization that the project required. Each month, approximately 200 hours of video and 200 hours of audio were digitized and encoded. The choice of digital data formats is essential and has been selected very carefully. Regarding the video recordings, the digitized content is stored in both uncompressed and lightly compressed (MPEG-2, 50 Mbit/s) formats. In the future, enhancements will always result in additional copies of the same content in the archives. The overall process follows the main steps described below:

1. Preparation of tape batches in Caux (same media format),
2. Scan in Vectracom (EPFL for HD content),
3. Visual and auditory quality control of digital media, input of results in the database,
4. Storage of compressed video and uncompressed audio on Amplidata hard disk system.

The quality control of digital archives is an extremely important operation. It is performed at EPFL, in part by students on campus, and in part by qualified personnel trained in the areas of audio or video. Mainly, this control is to determine whether the scan was performed to the required expectations:

- Audio levels: maximum peak and RMS average,
- The phase correlation and the audio balance,
- The shape of the spectrum,
- The maximum luminance levels (typically in the presence of spots) and minimum (black peeling),
- The color levels and color phase,
- Special kinds of audio and video defects.

This information is also of great value for research. The annotations performed by human operators serve as ground truth for algorithms or automated process control. After the quality control process has taken place, all information regarding defects are entered into a database.

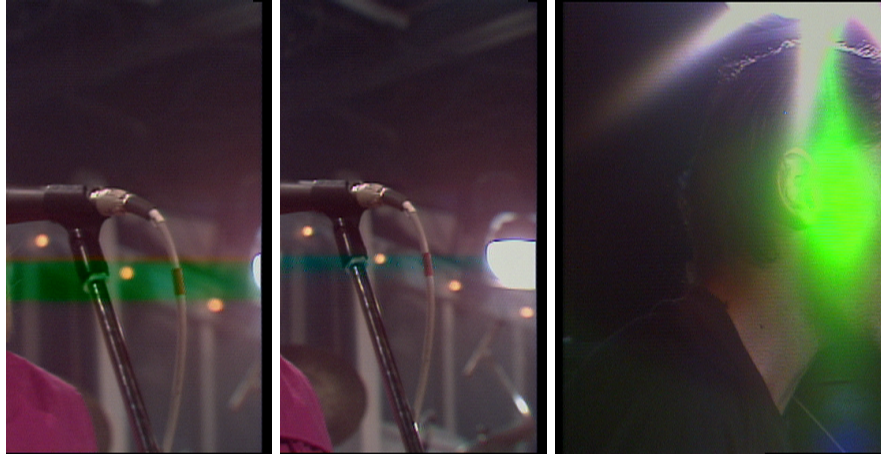


Figure 1: Bright light spots causing sensor burn in.

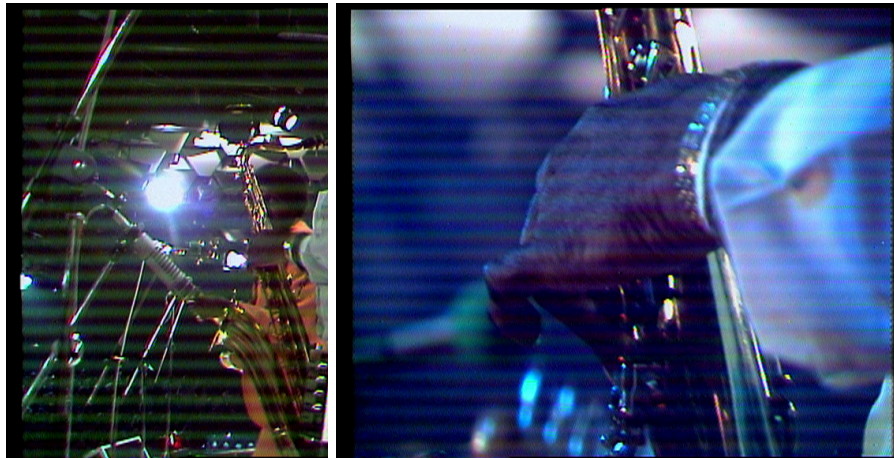


Figure 2: Proximity to a loud audio source causing modulation of the video signal by the audio signal.

2.2 Defects

A set of typical defects was established based on a detailed analysis of the digitized content. On the video quality part, only one typical defect is due to the digitalization process, whereas most of the defects come either from the filming conditions, the videotape recorder/player, or the aging of the tape. In the first case, the defects were already visible at the time of the recording and when the concerts were broadcast after the festival. Figure 1 illustrates one example of such defect. This defect is due to the presence of a bright light sources in the scene, such as a spotlight, and causes temporary sensor burn in. As a result, a greenish line or spot is visible in the video sequence. The intensity, size, and color of this effect typically change over the time (see Fig. 1). Figure 2 illustrates another typical defect due to filming conditions. Multiple medium-sized horizontal blue or green lines, sometimes moving up or down, characterize this defect. We have observed that this defect occurred when the camera operator was close to a very loud audio source (loudspeaker). This defect is mostly visible in the archives of the 70s and 80s. The origin of this defect is still not perfectly understood, but we suspect that the video signal was modulated by the audio signal due to the strong magnetic field generated by the loudspeaker cables and the close proximity of the camera cables and loudspeaker cables.

Different defects are also due to the videotape recorder and player. In particular, if the recording heads were not properly cleaned before the recording of a concert, the dirt on the heads would produce visible moving aperiodic horizontal lines on the video sequence. Figure 3 illustrates this defect. Another defect comes from the



Figure 3: Horizontal lines due to dirty heads on videotape recorder.

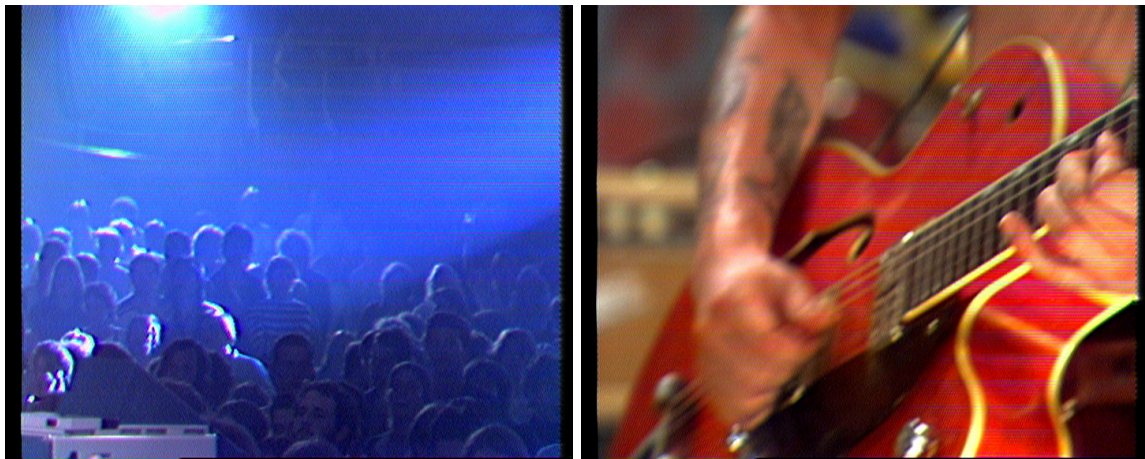


Figure 4: Cyclic groups of horizontal colored thin lines due to different tape-tape heads alignment between recorder and player.

vertical alignment between the tape and the heads. If the alignment on the videotape player differs significantly from the alignment on the videotape recorder, which can occur if a different machine is used for playback as opposed to recording, then cyclic groups of horizontal colored thin lines are visible on the video sequence. The video component is particularly sensitive to this misalignment due to the helical scan method. Figure 4 illustrates the defect visible on the video sequence. The period of the groups is related with the number of tracks per frame. We have observed that this defect is mostly visible on red and blue areas, and stronger on the right side of the frame (see Fig. 4). Ideally, the same machine should be used for both recording and playback. However, this condition is often impossible to fulfill considering that most of obsolete equipment is usually recycled.

It is well known that some videotape formats have a better longevity than others. In particular, the U-Matic, 2-inch, and 1-inch formats, which have mostly been used during the 70s and 80s, degrade quite fast with time. If the video heads fail to contact firmly with the tape because of roughness, wrinkling, particles or debris on tape or machine, then the video heads skip over the tape and the information cannot be read. If the magnetic oxide or coating flakes off, the information at that point is permanently lost. These losses of information are commonly referred to as dropouts. In the image, dropouts appear as pixels corresponding to missing data, black or white spots, or streak. The intensity of the defect typically decreases on the right part of the streak. On color content, the left part of the streak is sometimes composed of a few pixels with wrong color values. Figure 5 illustrates

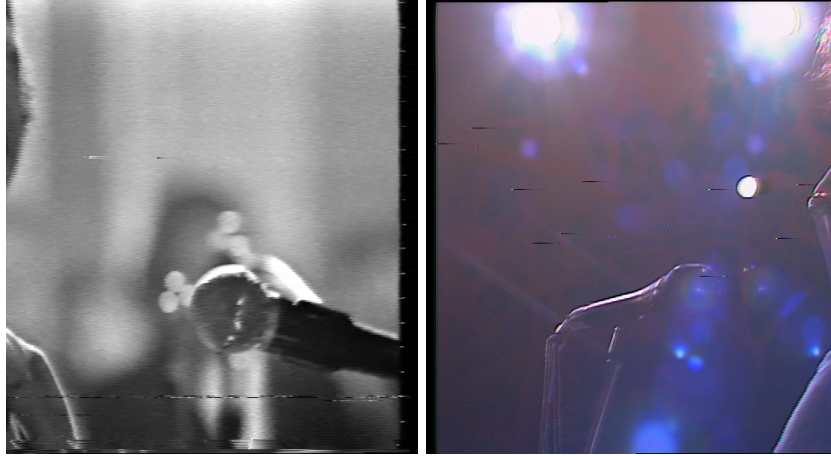


Figure 5: Dropouts due to flaked-off or rough tape's magnetic oxide or coating.

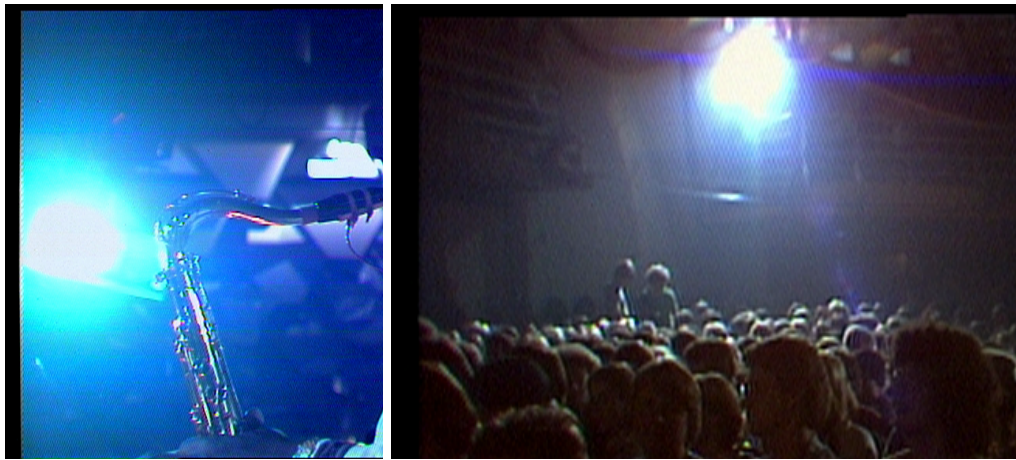


Figure 6: Diagonal grid due to jitter between the analog and sampling signals in the digitalization process.

dropouts on both black-and-white and color content. If several dropouts appear on each frame, then the video sequence appears snowy. Some videotape players are equipped with electronic compensators to substitute the missing information with information from the previous line. However, these systems are not perfect and some dropouts remain visible.

Finally, one typical defect was identified as resulting from the digitalization process. Figure 6 illustrates this defect. The origin of this defect comes from jitter between the original analog video signal and the sampling signal, which is used as synchronization for the digitalization process, due to imperfect calibration of the equipment. As a result, thin diagonal lines, with different periods and orientations, forming a grid, are visible on digitized frames. The grid pattern is more visible in some areas, depending on the texture, color, and luminance (see Fig. 6). This defect can be corrected by performing a proper calibration to ensure a perfect synchronization between the two signals. Therefore, the sequences that suffer from this defect are digitized again.

3. AUTOMATIC DEFECT DETECTION

Based on the analysis of the typical defects observed in the digitized content (see Sec. 2.2), their frequency of occurrence and impact on perceived quality, it was decided to first focus our efforts on two major defects: the audio lines and dropouts. This section describes the algorithms developed to automatically detect these two types of defect.

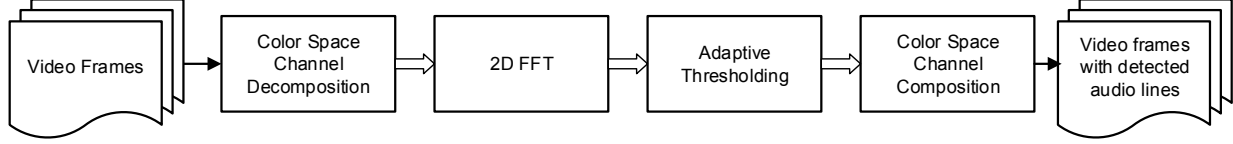


Figure 7: The block diagram of the algorithm for automatic detection of audio lines.

3.1 Audio lines

Structured, quasi-periodic noise, referred to as audio lines in this paper, appear as horizontal or slightly tilted periodic colored bands changing their amplitude, frequency, and phase frame-by-frame. Whereas such noise is spatially stationary in each frame, it introduces non-stationary properties in the temporal domain. For example, spatial periods of audio lines vary between 10 and 35 pixels within the investigated video dataset. Therefore, an automatic detection of particular frames containing this artifact together with precise determination of noise frequency needs to be performed *a priori* to the image enhancement phase. A relatively fast algorithm for automatic detection of audio lines in video sequences is described further in this section.

Figure 7 illustrates the block diagram of the proposed algorithm. Automatic detection is performed in each channel of used color space defined by the original format of video sequences (YUV 422 10 bit), and it is based on adaptive thresholding of Fourier amplitude spectrum.

Time variable audio lines pattern in spatial domain corresponds to peaks in Fourier spectrum varying mainly in their magnitude and location. Detection of these peaks is based on computation of adaptive threshold value for each pixel position $T_{i,j}$ using statistical properties of Fourier spectrum, namely mean value $M_{i,j}$ and standard deviation value $S_{i,j}$. To generalize this approach, the proposed algorithm considers arbitrary windowing elements $W_{M_{x,y}}$ and $W_{S_{x,y}}$ centered on position (i, j) . The mean and standard deviation are determined as

$$M_{i,j} = \text{mean} \left(\bigcup_{x,y} W_{M_{x,y}}^{i,j} A_{x,y} \right) \quad (1)$$

$$S_{i,j} = \text{std} \left(\bigcup_{x,y} W_{S_{x,y}}^{i,j} A_{x,y} \right) \quad (2)$$

Then, the threshold value is given as a linear combination of the above values as

$$T_{i,j}^{W_M W_S} = C_M M_{i,j} + C_S S_{i,j} \quad (3)$$

where C_M and C_S are constant multipliers defining the threshold value. The detection is successful for all amplitude spectral components for which $A_{i,j} > T_{i,j}^{W_M W_S}$.

This general approach is applicable to any periodic and quasi-periodic noise within the entire frame. Since we are focused on a particular artifact, a spatial spectral mask is applied choosing only spectral components corresponding to properties of audio lines. Range of values in audio lines periods defines the radial limits for a spatial spectral mask in shape of annulus whereas their inclination determines the angular sector.

The threshold value depends on four different parameters: width of windowing elements (W_M and W_S) and values of constant multipliers (C_M and C_S). Choice of the width of a window is strongly connected to the size of the peaks. To detect a particular frequency, the width of the window must be bigger than the emphasized area around the peak. The constant multipliers define the sensitivity of thresholding. For the dataset under evaluation, the window size 15×15 pixels was empirically found as an optimal choice, and several different values for constant multipliers were tested to obtain optimal thresholding value (see Sec 4.1).

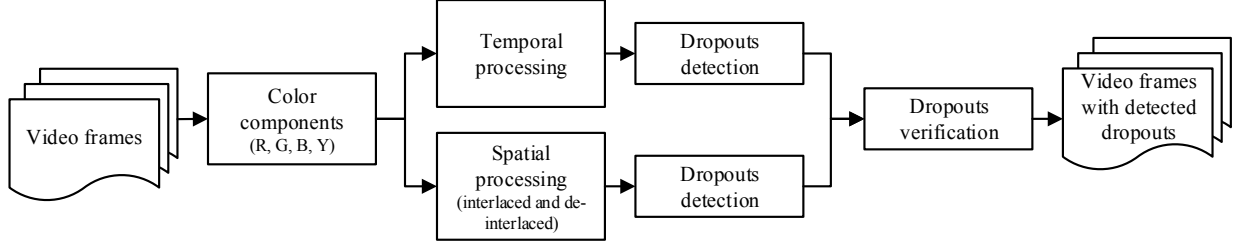


Figure 8: The block diagram of the algorithm for automatic detection of dropouts.

3.2 Dropouts

Dropouts are local defects in video sequences and appear as horizontal segments with significantly different pixel values from neighboring pixels (see Sec. 2.2). Therefore, it is important to detect only those local regions in video frames that are distorted and that will be removed by further filtering. In this section, we present our method for automatic detection of dropouts in video sequences.

In the Montreux Jazz Festival Digital Archives, dropouts appear in video frames in a highly irregular manner, i.e., they are not correlated between a given frame and its next frame. Therefore, for reliable dropouts detection, we suggest exploiting both spatial and temporal nature of frames. Figure 8 illustrates the block diagram of the detection method. Every frame is represented in both RGB and YCbCr color spaces. The R, G, B, and Y components are separately considered for further processing, as explained in details in the following.

First, the temporal nature of the frames is considered for dropouts detection. A reasonable assumption is made that dropouts do not appear at the same positions in consecutive frames, based on observations of the dropouts temporal nature in the digital archives. To detect pixels as potential dropouts, we introduce the Dropout Detection Index (DDI), which gives a useful measure of the presence or absence of any local distortion in a video frame:

$$DDI^c = \begin{cases} \frac{|d_p + d_n|}{1 + |d_p - d_n|}, & \text{for } d_p > K \cdot \text{mean}_{local}(d_p) \text{ and } d_n > K \cdot \text{mean}_{local}(d_n) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where c , p , and n refer to the current, previous, and next frame, respectively, and $d_p = |c - p|$ and $d_n = |c - n|$ are the differences between the current and previous frames, as well as between the current and next frames. Parameter K was empirically determined and set to 2. DDI^c is computed for each pixel, therefore it is of the same dimension as the resolution of the frames c in a video sequence. Instead of thresholding DDI^c with a constant value across the frame, a multiple of the local mean of the differences d_p and d_n is used as adaptive threshold. The local means are calculated for each block of some fixed size (5×5 pixels, in our calculations). In effect, this makes the thresholding operation adaptive to the local spatial information and local movement, thus improving the detector performance.¹⁴ In addition, to decrease false detection, only blocks where motion in both d_p and d_n is low are considered. Low motion is determined by making use of hysteresis thresholding with the lower and upper threshold values set to 25 and 40, respectively. As a final result of this stage, $DDI_{temporal}^c$ is determined for every frame c .

Second, the spatial nature of the frames is considered for dropouts detection. Here we assume that dropouts do not appear in consecutive rows and that horizontal edges in frames are thicker than 1 pixel. However, due to interlacing, these assumptions can lead to false detection of horizontal edges when there is a large vertical motion. Therefore, we consider detection of dropouts in the de-interlaced versions of the frame, namely the two fields are separated by splitting the even and odd rows of the original frame and both fields are separately processed. DDI^c is now computed with Eq. (4) in an analogous way, by comparing the current, previous, and next rows within one field. Three different DDI^c measures are calculated separately for the interlaced signal and the two fields of the same frame c . Two DDI^c measures corresponding to the two fields are at the end combined into a single $DDI_{spatial_deinterlace}^c$ by performing the dual operation of the deinterlacing process. As a final result of this stage, $DDI_{spatial_deinterlaced}^c$ and $DDI_{spatial_interlaced}^c$ are determined for every frame c .

Then, the three measures computed in the previous stages, namely $DDI_{temporal}^c$, $DDI_{spatial_interlaced}^c$, and $DDI_{spatial_deinterlaced}^c$, are processed to determine the relevant dropouts and discard the false ones. Remind that these measures are of the same dimension as the resolution of a frame c . Since dropouts appear as horizontal segments, the values along each row in DDI^c are first summed to determine the rows with the highest intensity change. We reasonably assume here that the higher intensity change in a row, the higher the probability of dropouts presence in that row. Only N rows with the highest values are selected. The parameter N was empirically set to 10. Then, the adaptive threshold of $\mu + 2\sigma$ is applied on the selected rows, where μ and σ are the mean and standard deviation of DDI^c , respectively. As a result of this phase, discontinuities that appear in the selected rows with a total intensity change higher than the threshold are selected as dropouts. These calculations are repeated for all three DDI^c measures, and three binary masks with locations of dropouts are determined. Morphological filtering (dilation) is further applied on binary masks to include possibly missing dropouts in every line.

Finally, the results represented by the three binary masks, i.e., temporal, spatial interlaced, and spatial deinterlaced, are combined to determine unique and valid dropouts. The decision on the final dropouts is made as follows: each dropout should be detected in at least two binary masks and in at least two components (R, G, B, or Y). To further reduce the false detection, we suppose that frame filtering in the regions of detected dropouts modifies the perceived frame. Therefore, we check if the improvement by applying a simple median filtering in the regions of detected dropouts (or the degradation, in case of false detection) is visible by means of the structural similarity (SSIM) map. In filtering, every pixel detected as dropout is replaced by the median value of its seven neighboring pixels from the previous and the next rows, and from the left column. The SSIM map is then computed between the original and filtered frames. Finally, pixels are classified as visible (valid dropouts) and non-visible dropouts by making use of hysteresis thresholding applied on the SSIM map, with the lower and upper threshold values set to 0.2 and 0.65, respectively. As a final outcome, visible dropouts that have an impact on the perceived image are kept as valid dropouts.

Pixels detected as valid dropouts in the original frame are replaced by the corresponding pixels from the filtered frame. This updated frame is used as a new input for a second pass of the algorithm to allow dropouts with lower confidence to be tested. Valid dropouts detected during each pass are merged to produce the final result of the detection process.

4. RESULTS

A list of the most critical defects was built by the MetaMedia Center (see Sec. 2) to identify both the more perceptually annoying and widely spread defects. During quality control operations, concerts showing significant defects were tracked and kept apart for analysis. Then, for the purpose of automatic defect detection and enhancement, EPFL students were involved in indexing and tagging those critical video sequences to precisely indicate when and where the defects appear on the video streams. Many samples of each defect were extracted to build a ground truth database to benchmark the algorithms. All video sequences were available in standard definition, 25 fps, interlaced format. The database contained a total of 28 and 9 different video sequences for the evaluation of the audio lines and dropouts detection algorithms, respectively. Tables 1 and 2 detail various statistics on video sequences used in the ground truth database.

To evaluate the performance of the developed algorithms, standard receiver operating characteristics (ROC) were used. Given the ground truth and the detection results, the numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were computed for each video sequence. Based on these numbers, the following set of complementary performance metrics was computed to evaluate the proposed algorithms: true positive rate (TPR), false positive rate (FPR), accuracy (ACC), specificity (SPC, or true negative rate), positive predictive value (PPV), negative predictive value (NPV), and false discovery rate (FDR).¹⁵ The following subsections report the results for the audio lines and dropouts detection.

Table 1: Statistics of the ground truth database used to evaluate the audio lines detection algorithm.

Video	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Number of frames	252	111	144	451	134	396	225	148	130	98	63	178	102	54	210	66	149	77	60	35	69	46	38	73	168	77	35	81
Number of frames with audio lines	45	76	84	154	43	146	77	69	98	24	18	26	59	32	29	14	101	37	60	22	33	11	28	68	79	62	15	64

Table 2: Statistics of the ground truth database used to evaluate the dropouts detection algorithm.

Video	1	2	3	4	5	6	7	8	9
Number of frames	15	43	29	5	21	161	13	24	25
Number of frames with dropouts	10	42	0	1	6	45	2	0	7
Number of dropouts	15	230	0	1	8	60	2	0	49
Average number of dropouts per frame	1	5.3	0	0.2	0.4	0.4	0.2	0	2.0
Max number of dropouts per frame	3	14	0	1	2	4	1	0	21

4.1 Audio lines

Figure 9 shows a typical example of image data, its Fourier amplitude spectrum (Cb channel), used spatial spectral mask, and results for detection of horizontal audio lines in each color channel. All peaks corresponding to quasi-periodic noise are shown and detected, even for higher frequencies. Applying the mask (see Fig. 9(c)), the result of detection can be narrowed to audio lines corresponding frequencies only.

The goal was to find one combination of parameters to automatically detect audio lines with high accuracy through the entire dataset. Four variables mentioned above, i.e., two widths of windowing elements for computing mean and standard deviation and two constant multipliers were changed to achieve the optimal combination. The window width of 15×15 pixels was determined as optimal for SD resolution content and Fourier amplitude spectrum. Several combinations of C_M and C_S were evaluated to determine optimal thresholding parameters for the entire video dataset. Standard ROC were used to evaluate particular combinations of parameters. To choose the optimal combination of C_M and C_S , the overall accuracy was taken as a decision-making criterion. Figure 10 shows the ROC curves expressing the relation between TPR and FPR for the following parameters: $W_M = W_S = 15 \times 15$ pixels, $C_M = 2.5$, and $C_S = [2, 3]$. For better orientation, only the set of ten most representative video sequences is shown. One can see that detection performance depends strongly on the threshold value. High values of C_S result in poor detection performance and TPR is at its lower limit for each content. On the other hand, there are many false positive detections for low values of C_S so we get a lot of false alarms detecting clear frames. Consistent trend of all characteristics shows that a proper combination of parameters can improve the detection in order of tens of percent. The goal is to choose the optimal combination of C_M and C_S that will provide best detection performance for all frames over the entire dataset, namely that TPR and FPR will be as close as possible to 1 and 0, respectively. Based on the different ROC curves, the parameters $C_M = C_S = 2.5$ were chosen to complete the optimal set of parameters.

Table 3 reports the ROC for the parameters $W_M = W_S = 15 \times 15$ pixels and $C_M = C_S = 2.5$. From overall point of view, the detection algorithm shows a good performance within the evaluated dataset with average value of TPR = 87%, ACC = 83%, and PPV = 76%. The biggest deviation from the average performance in terms of accuracy can be seen for video sequences 7 and 16 with accuracy values of 67% and 62%, respectively. This is caused by higher values of false positive (21% and 38%, respectively) in relation of number of frames. Closer analysis of amplitude spectrum of problematic frames shows that defined thresholding parameters are too sensitive in comparison to given ground truth. In another words, even if the artifact doesn't disturb the viewer, it is still detected in spectrum. On the other hand, sequences 8 and 10 show higher values of false negative, meaning that the detection algorithm missed frames for which the positive ground truth is defined. For further processing, which is, as mentioned above, the removal of the artifact, it is crucial to lower false alarms and missed detections to zero. Therefore, the thresholding parameter should be adapted and tailored to the content.

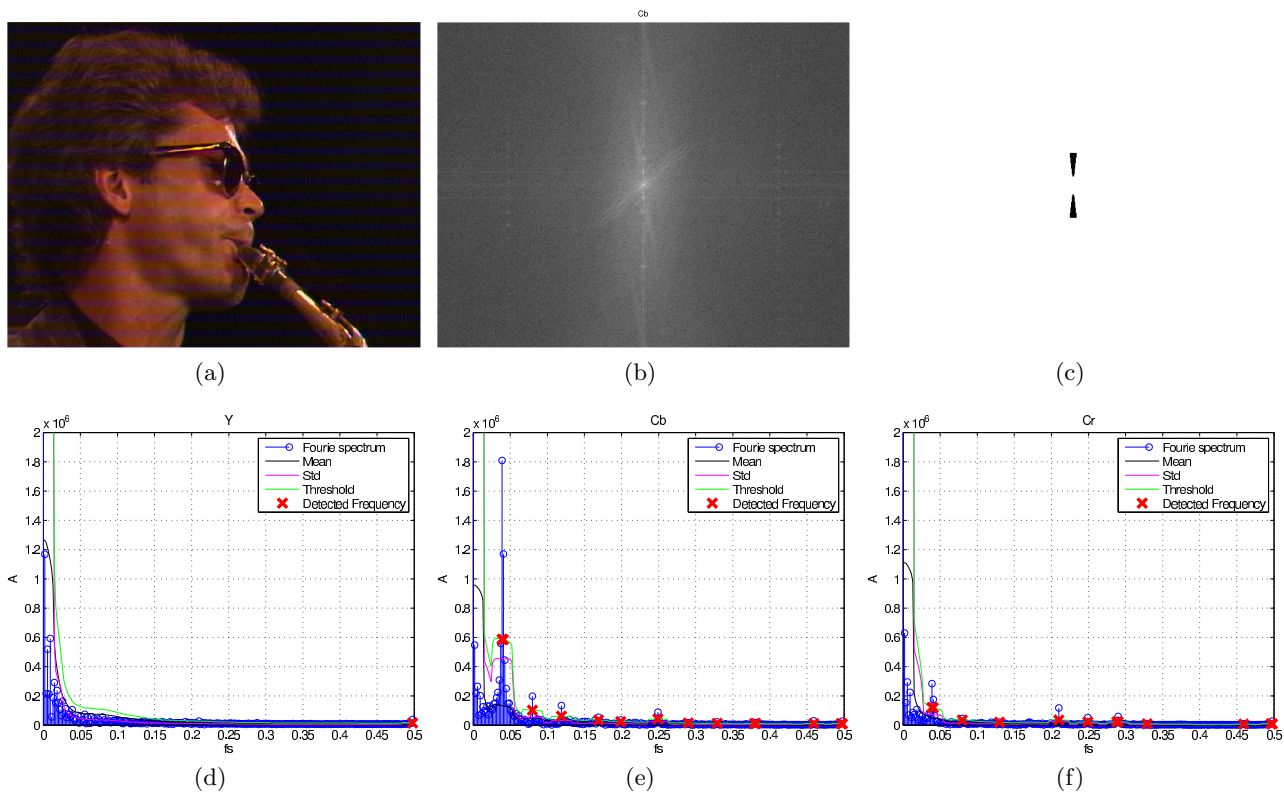


Figure 9: a) Typical image data containing the audio lines. b) Fourier amplitude spectrum of Cb component. c) Spatial spectral mask used for detection. d-f) results for detection of horizontal lines before applying spectral mask.

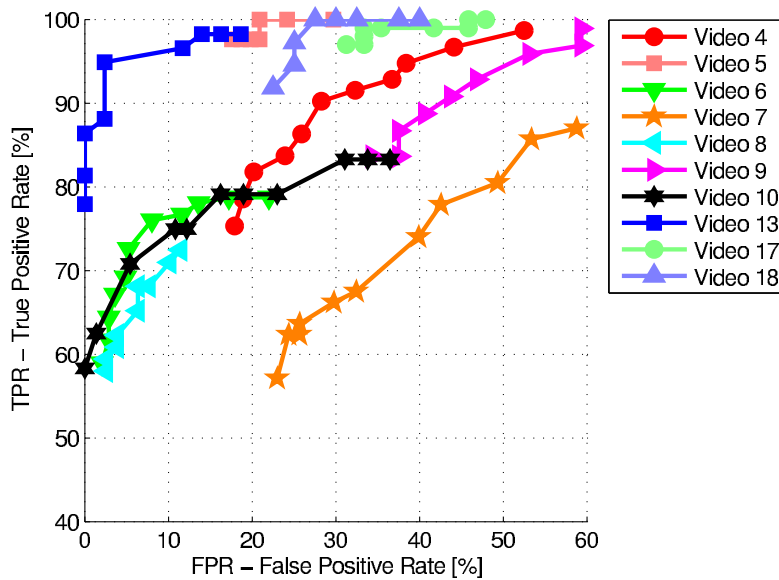


Figure 10: Receiver operating characteristics of the proposed algorithm for audio lines detection.

Table 3: The performance metrics for audio lines detection. All values are given in [%].

	TP	TN	FP	FN	TPR	FPR	ACC	SPC	PPV	NPV	FDR
Video 1	6.35	78.57	3.57	11.51	35.56	4.35	84.92	95.65	64	87.22	36
Video 2	54.05	27.03	4.5	14.41	78.95	14.29	81.08	85.71	92.31	65.22	7.69
Video 3	53.47	40.28	1.39	4.86	91.67	3.33	93.75	96.67	97.47	89.23	2.53
Video 4	30.82	47.23	18.63	3.33	90.26	28.28	78.05	71.72	62.33	93.42	37.67
Video 5	31.34	53.73	14.18	0.75	97.67	20.88	85.07	79.12	68.85	98.63	31.15
Video 6	26.77	59.85	3.28	10.1	72.6	5.2	86.62	94.8	89.08	85.56	10.92
Video 7	23.11	44.44	21.33	11.11	67.53	32.43	67.56	67.57	52	80	48
Video 8	29.05	51.35	2.03	17.57	62.32	3.8	80.41	96.2	93.48	74.51	6.52
Video 9	68.46	13.85	10.77	6.92	90.82	43.75	82.31	56.25	86.41	66.67	13.59
Video 10	19.39	63.27	12.24	5.1	79.17	16.22	82.65	83.78	61.29	92.54	38.71
Video 11	20.63	71.43	0	7.94	72.22	0	92.06	100	100	90	0
Video 12	4.49	75.84	9.55	10.11	30.77	11.18	80.34	88.82	32	88.24	68
Video 13	54.9	41.18	0.98	2.94	94.92	2.33	96.08	97.67	98.25	93.33	1.75
Video 14	57.41	29.63	11.11	1.85	96.88	27.27	87.04	72.73	83.78	94.12	16.22
Video 15	13.81	57.62	28.57	0	100	33.15	71.43	66.85	32.58	100	67.42
Video 16	21.21	40.91	37.88	0	100	48.08	62.12	51.92	35.9	100	64.1
Video 17	67.11	20.81	11.41	0.67	99.01	35.42	87.92	64.58	85.47	96.88	14.53
Video 18	48.05	36.36	15.58	0	100	30	84.42	70	75.51	100	24.49
Video 19	96.67	0	0	3.33	96.67	0	96.67	0	100	0	0
Video 20	57.14	20	17.14	5.71	90.91	46.15	77.14	53.85	76.92	77.78	23.08
Video 21	47.83	24.64	27.54	0	100	52.78	72.46	47.22	63.46	100	36.54
Video 22	23.91	47.83	28.26	0	100	37.14	71.74	62.86	45.83	100	54.17
Video 23	73.68	18.42	7.89	0	100	30	92.11	70	90.32	100	9.68
Video 24	91.78	1.37	5.48	1.37	98.53	80	93.15	20	94.37	50	5.63
Video 25	47.02	30.95	22.02	0	100	41.57	77.98	58.43	68.1	100	31.9
Video 26	79.22	14.29	5.19	1.3	98.39	26.67	93.51	73.33	93.85	91.67	6.15
Video 27	40	51.43	5.71	2.86	93.33	10	91.43	90	87.5	94.74	12.5
Video 28	79.01	7.41	13.58	0	100	64.71	86.42	35.29	85.33	100	14.67
Average	45.24	38.2	12.14	4.42	87.08	26.75	83.44	69.68	75.59	86.06	24.41

Table 4: The performance metrics for dropouts detection. All values are given in [%].

	TP	TN	FP	FN	TPR	FPR	ACC	SPC	PPV	NPV	FDR
Video 1	0.16	99.51	0.32	0.01	93.33	0.32	99.67	99.68	34.15	99.99	65.85
Video 2	0.72	98.31	0.75	0.22	76.52	0.76	99.03	99.24	48.89	99.78	51.11
Video 3	0.00	99.66	0.34	0.00	100.00	0.34	99.66	99.66	0.00	100.00	100.00
Video 4	0.04	99.82	0.14	0.00	100.00	0.14	99.86	99.86	20.00	100.00	80.00
Video 5	0.03	99.84	0.09	0.03	50.00	0.09	99.87	99.91	26.67	99.97	73.33
Video 6	0.05	99.56	0.37	0.01	83.33	0.37	99.62	99.63	12.79	99.99	87.21
Video 7	0.03	99.81	0.16	0.00	100.00	0.16	99.84	99.84	14.29	100.00	85.71
Video 8	0.00	99.74	0.26	0.00	100.00	0.26	99.74	99.74	0.00	100.00	100.00
Video 9	0.29	99.12	0.54	0.05	85.71	0.54	99.41	99.46	35.29	99.95	64.71
Average	0.15	99.49	0.33	0.04	87.66	0.33	99.63	99.67	21.34	99.96	78.66

4.2 Dropouts

The proposed algorithm determines lines in video frames where dropouts appear, therefore it can be evaluated as a typical detection task. The goal of this detection task is to classify every line in a video frame as a line with or without dropouts. Table 4 reports the values of performance metrics for every video separately, as well as the average values for the entire video dataset. Values are calculated for the set of parameters which are empirically selected (see Sec. 3.2). In average, the algorithm achieves good performance in terms of high TPR, high ACC, and low FPR. Video sequences with few dropouts show good TPR (above 85%) as the majority of dropouts are detected by the proposed algorithm. For video sequences with higher density, TPR decreases to 75% or below. High values of ACC, SPC, and NPV are expected due to very high number of true negatives, i.e., high number of lines where there are no dropouts. Also, the number of false positives is limited to 20 lines per frame by a parameter, which defines the number of the rows with the highest intensity change (2 passes with $N = 10$ lines each, see Sec. 3.2), whereas the minimum number of lines with true negatives is around 550. Therefore, the value of FPR is kept very low - in worst case it can reach 0.76%. As complementary metrics to those presented, SPC and NPV have lower limit of 99.2%. On the other hand, in average, the algorithm introduces four times more false positives than the number of true positives in a video, as shown by FDR and PPV values. The false positives are introduced in regions with high motion, where pixel differences in consecutive frames tend to be high. Therefore, additional processing step can be introduced to further reduce false positives. For example, determining the displacement of the blocks in consecutive frames and calculating differences between local, displaced blocks, instead of computing differences over an entire frame.

Figure 11 shows the performance of dropouts detection in the form of the ROC curves. The curves are shown for the set of video sequences containing dropouts (see Tab. 2). The curves were produced by varying parameter K in Eq. (4) between 1 and 20. Results show that TPR and FPR are highly dependent on the selection of the parameter K . General trend is that both TPR and FPR increase by decreasing the value K , except for K values between 1 and 3 where, for majority of the video sequences, the TPR becomes “saturated” once it reaches the highest extreme values. It is interesting that these extreme values are different for different video sequences, and that the proposed automatic approach performs better for sequences with lower number of dropouts. For instance, the TPR in sequences 1, 4, 5, and 7 with few dropouts achieve their highest values of 100%, when compared to sequences 2, 6, and 9 with the high number of dropouts where the TPR reaches upper limit around 80%. Also, sequences with smaller average number of dropouts per frame show better performance in term of FPR. For instance, video sequence 5 reaches its best performance of TPR = 100% for FPR = 0.1%, whereas sequence 2 shows best performance for FPR = 0.7%. Another interesting observation is that video sequences do not reach their best performance for the same value of K , and that the same variations of K produce different changes in performance of different sequences. For example, video sequences 1, 5, and 7 reach 100% of TPR quite fast and then keep this value while FPR increases. On the other hand, video sequence 6 is slow in reaching its higher TPR values. Taking into account a tradeoff between high TPR and low FPR and their different values for every video, we selected the value 2 for K as the optimal parameter for all video sequences.

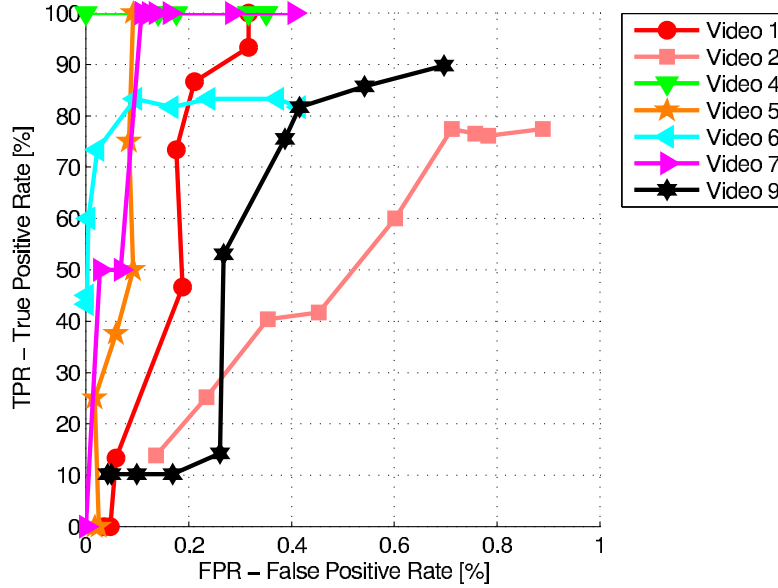


Figure 11: Receiver operating characteristic of the proposed algorithm for dropouts detection.

These results show that the proposed algorithm performs very well for sequences with lower number of dropouts, whereas performance decreases for sequences with a lot of dropouts. This and all the other facts presented in this section lead to a conclusion that parameters should ideally be adjusted for each sequence separately to achieve higher performance for each of them. One of the solutions is to manually select values for different video sequences, which would lead to a semi-automatic approach. However, having large number or long video sequences in a digital archive would require a time-consuming and tedious process of parameter tuning. Therefore, we suggest at first extending the objective analysis presented in this paper to include a subjective assessment of restored video sequences, i.e., where the detected dropouts are automatically removed by filtering. Subjective assessment can determine the lower limit of performance that would meet subjects' expectations for different content types. By doing this, we can potentially reduce the time for parameter tuning and faster achieve satisfactory results.

5. CONCLUSION

In this paper, we presented two algorithms to automatically detect multiple quasi-periodic horizontal lines, referred to as audio lines, and dropouts, which are two typical defects visible in video archives. Audio lines correspond to peaks in the Fourier spectrum and were detected based on an adaptive thresholding of the frequency components. Dropouts appear as pixels corresponding to missing data and were detected by comparing each pixel to its spatial and temporal neighbors and applying an adaptive thresholding. These algorithms were applied to the Montreux Jazz Festival Digital Archives. To evaluate the performance of the algorithms, a ground truth database was created by collecting annotations from human subjects on a set of video sequences showing these two defects. Results show that both algorithms achieve good performance with a true positive rate of above 87%.

To improve the performance of defect detection, thresholding parameters should ideally be adjusted for each sequence separately, for example by manual selection. However, such semi-automatic approach is not convenient when a large amount of data needs to be processed. For video archive restoration, defect detection is closely linked with defect removal. In future work, the performance of the developed algorithms will be evaluated through subjective evaluation of restored video sequences. Such evaluation will be very beneficial to determine the impact of false detection on perceived quality as well as an optimal tradeoff between false positive and false negative.

ACKNOWLEDGMENTS

This work has been conducted in the framework of the Swiss National Foundation for Scientific Research (FN 200021-143696-1), EC funded Network of Excellence VideoSense, and COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services QUALINET.

REFERENCES

- [1] Richardson, P. and Suter, D., “Restoration of historic film for digital compression: a case study,” in [*IEEE Int. Conf. on Image Processing (ICIP)*], **2**, 49–52 (Oct. 1995).
- [2] Chenot, J., Drewery, J., and Lyon, D., “Restoration of archived television programmes for digital broadcasting,” in [*Proc. IBC*], 26–31 (1998).
- [3] van Roosmalen, P. M. B., [*Restoration of archived film and video*], Universal Press (1999).
- [4] Hamid, M., Harvey, N., and Marshall, S., “Genetic algorithm optimization of multidimensional grayscale soft morphological filters with applications in film archive restoration,” *IEEE Trans. on Circuits and Systems for Video Technology* **13**, 406–416 (May 2003).
- [5] Gonzalez, R. and Woods, R., [*Digital Image Processing*], Pearson/Prentice Hall (2008).
- [6] Schowengerdt, R., [*Remote Sensing: Models and Methods for Image Processing*], Elsevier Science (2006).
- [7] Aizenberg, I. N. and Butakoff, C., “Frequency domain medianlike filter for periodic and quasi-periodic noise removal,” (2002).
- [8] Crow, F. C., “Summed-area tables for texture mapping,” *SIGGRAPH Comput. Graph.* **18**, 207–212 (Jan. 1984).
- [9] Abbas, J. K. and Domanski, M., “Rejection of scratches from video by use of filters with prediction error processing,” *Proc. of Signal Processing X: Theories and Applications, EUSIPCO* (2000).
- [10] Harvey, N. and Marshall, S., “Defect detection from scratched 2-inch quadruplex video tape,” in [*7th Int. Conf. on Image Processing And Its Applications*], **1**, 465–468 (July 1999).
- [11] Kaprykowsky, H., Liu, M., and Ndjiki-Nya, P., “Restoration of digitized video sequences: An efficient drop-out detection and removal framework,” in [*16th IEEE Int. Conf. on Image Processing (ICIP)*], 85–88 (Nov. 2009).
- [12] Joyeux, L., Buisson, O., Besserer, B., and Boukir, S., “Detection and removal of line scratches in motion picture films,” in [*IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*], 553–558 (June 1999).
- [13] Shih, T., Lin, L., and Lee, W., “Detection and Removal of Long Scratch Lines in Aged Films,” in [*IEEE Int. Conf. on Multimedia and Expo (ICME)*], 477–480 (July 2006).
- [14] Kokaram, A. C., *Motion picture restoration*, PhD thesis, University of Cambridge (1993).
- [15] Fawcett, T., “An introduction to ROC analysis,” *Pattern Recognition Letters* **27**, 861–874 (June 2006).