EFFICIENT RESTORATION OF VARIABLE AREA SOUNDTRACKS

Abdelâali Hassaïne[1], Etienne Decencière[1] and Bernard Besserer[2]

[1]Centre de Morphologie Mathématique, Ecole Nationale Supérieure des Mines de Paris, 77305 Fontainebleau, France, [2]Laboratoire Informatique, Image, Interaction, Université de La Rochelle, 17042 La Rochelle, France
e-mail: {hassaine,decencie}@cmm.ensmp.fr, bernard.besserer@univ-lr.fr

ABSTRACT

The restoration of motion picture films using digital image processing has been an active research field for many years. The restoration of the soundtrack however, has mainly been performed in the sound domain, using signal processing methods, in spite of the fact that it is recorded as a continuous image between the images of the film and the perforations.
In this paper a restoration method for variable area soundtrack restoration at the image level is presented. First, a novel method is proposed for the detection of the symmetry axis of the scanned soundtrack. Then, a comparison between the watershed and the region growing segmentation of the soundtrack is developed. Another algorithm aiming to enforce the symmetry and to correct the edges of the segmented image is presented. A last step aiming to smooth the edges of the obtained image is performed. Finally, experimental results are reported and possible future improvements are discussed.

Keywords: image processing, region growing, symmetry enforcement, variable area soundtrack restoration.

INTRODUCTION

The majority of the films shot since the Thirties use an optical soundtrack to record the sound. The standard Academy Optical Mono track was introduced by the Academy of Motion Picture Arts and Sciences. According to this standard, the optical soundtrack is located in a space of 3 mm between the images and the perforations. Thus, the duplication of a film makes it possible to copy its soundtrack. In its classical version of area variation, the optical soundtrack contains a central bright region and two lateral dark regions. The width of the white region is directly proportional to the amplitude of the audio signal.

Unfortunately, the optical soundtrack undergoes the same type of degradations as the image of the film (*e.g.*, scratches, dust), but as the soundtrack is close to the film edge and perforations, it is sometimes degraded by abrasion or impaired over a large surface due to moisture.

Typically, sound processing and restoration are performed only after the transformation of the optical information into acoustic electric signal. Unfortunately, several defects introduce distortions that are delicate to correct after the transformation. As powerful as they are, digital audio processing systems cannot make the difference between some audio artifacts caused by the degradation of the optical soundtrack, and some sounds present in the original soundtrack.

The existing methods that aim to restore the soundtrack by image-processing seem to be rudimentary and deal with a limited number of defects. We proposed in (Brun *et al.*, 2007) a method which gives generally good results, however when dealing with high frequencies images, it does not offer an accurate localization of edges. More generally, it does not provide a precise detection of thin structures. Moreover, it requires a long computation time.

The method we propose in this paper, overcomes the drawbacks of the previous one and uses new techniques for the restoration of variable area soundtracks. The expertise on optical soundtracks as well as the validation of the results are due to our collaboration with motion picture manufacturers, archivists and experts. Real-time processing is not a main concern for the moment, however a fast computation time will be an additional asset. The images to be treated and their defects are presented in the following section. The third section sums up the state of the art. Finally, in the fourth and fifth sections, the proposed method and its results are presented.

OPTICAL PICKUP AND IMAGE ACQUISITION

PRINCIPLES

The optical soundtrack is read by projecting light from a lamp through a narrow slit, then through

the film. A photodetector detects the transmitted intensity and converts the changes of brightness to an electrical modulation, further processed and amplified to generate the sound. The most commonly used standard is the fixed density process (also called variable area). It is still in use today. In this process, the varying proportion of black versus white modulates the intensity. This study only concerns this type of soundtracks (Fig. 1).
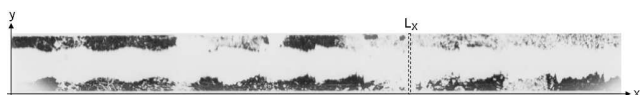

Fig. 1. *Damaged variable area / fixed density soundtrack.*

DIGITAL INPUT IMAGES

The so-called Academic 'A' chain components includes the optical pickup and the preamplifier (ISO, 1986). The defined bandwidth for these components is roughly 40 Hz-10 KHz (according to the Academic curve, used from 1930 until the fifties, and the introduction of the Dolby noise reduction systems). Since the film speed is 456 mm/s (for 35mm films), a 10 KHz signal period is 0.0456 mm wide in the optical soundtrack.

Two databases have been used in this study. The first one has been produced with a frame-by-frame scanner with a 1536 lines definition. According to a frame height of 16 mm (4 perf.), the image resolution is ca. 100 pixels/mm ($\approx$ 2500 DPI) and one period of a 10 KHz signal ranges over 4 pixels. This limited resolution was good enough for our first trial. In addition to those images, the images of the second database have been collected by a specific line-scan based digitizer able to sample 48 000 lines of 512 pixels per second with a resolution of 3200 DPI and a gray level depth of 8 bits, thus good enough for processing in high frequencies (meeting the requirements of the 'X'-curve (ISO, 1987)). The scanned soundtracks are taken from old French movies, they contain both music and voice. Several defects are present in these soundtracks including dust, scratches and molds.

STATE OF THE ART

Streule (1999) proposed a soundtrack restoration system using image processing tools. Concerning the restoration, Streule only treats defects caused by dust. This technique is mainly based on the soundtrack symmetry.

J. Valenzuela appears as the inventor of several patents on soundtrack scanning and restoration (see for instance United States patent 7,123,339). He proposes a short description of his technique in Valenzuela (2003). The restoration is very simple, and is based on median filters and erosions. It can only deal with small defects.

Richter *et al.* (2003) proposed a method for defect localization in multiple double sided variable area soundtracks. This method applies first a low-pass filter to the image. After binarization, the remaining defects are sufficiently large to be easily detected. The authors did not consider the problem of their elimination.

Spots detection has also been studied by Kuiper (2005). The spots being lighter than the other parts of the image, a threshold isolates them. A succession of morphological operations is then applied for a better spot localization and for the removal of isolated pixels. Unfortunately, in most cases, the spots are not lighter than the rest of the image. Therefore, this method cannot always be used.

As a conclusion, it seems that the research on this field is at its beginnings. The published approaches do not allow to treat the whole range of defects we have found in our database, or that we have identified through the expertise of our partners.

We proposed in Brun *et al.* (2007) a new restoration method which first filters the image using a morphological opening by reconstruction to reduce the defects in the initial image (Fig. 2b). Then, the quasi-distance of the image is computed (Beucher, 2007) (Fig. 2c). The symmetry axis is obtained through a linear regression of crest points of the quasi-distance image (Fig. 2d). The image is then segmented using a watershed on the morphological gradient of the image. The watershed is initialized with three markers: the symmetry axis is chosen as a central marker, the two other markers correspond to the two lines parallel to the axis. The distance between the central marker and the side ones is given by the maximum value of the quasi-distance, plus an empirical value (Figs. 2e and 2f). The final step enforces the symmetry of the obtained segmentation and corrects its edges by favoring the closest black pixel to the symmetry axis (Fig. 2g). We tested this method on the images of our first database and, according to motion picture archivists, the results were satisfactory. However, when we applied it to the second database, it appeared that thin structures, produced by high frequencies, were not correctly restored. Furthermore, this method requires a long processing time which is mainly due to the quasi-distance computation. Finally, when using this method, the audio signal corresponding to

the restored soundtrack contains a background noise which is more audible than in the original version. As we will see further in this paper, this is due to high frequencies introduced by the binarization process.

In this paper, we will improve and complete our previous method and make it more appropriate for the new database.

RESTORATION METHOD

OBJECTIVE IMAGE MODEL

An image f is modeled as an integer function of an appropriate subset of $\mathbb{Z}^2$. We define the *x-axis* and the *y-axis* which respectively correspond to the horizontal axis (*i.e.*, the direction of the film movement) and the vertical axis (Fig. 1).

We model the original soundtrack images as binary images containing a central white region and two black side regions. These images are symmetrical with respect to a central axis. The symmetry axis is not exactly horizontal, however its slope is very small ($\leq 2°$). This point will be taken into account in the determination of the symmetry axis. Moreover, the contours of the two side regions must be *causal*, *i.e.*, every x-coordinate is uniquely associated with one point of each contour (for instance, the contours are causal in Fig. 2h but in Fig. 2f they are not).

PROCESSING STEPS

The proposed restoration method consists of the following steps:

– Symmetry axis detection

– Image segmentation

– Symmetry enforcement and edge correction, and

– Contours smoothing

In our previous method, the image is first filtered by a morphological opening by reconstruction in order to reduce the defects in the initial image. We noticed that the image improvement due to this step is not important for the subsequent steps of our method and the final results do not change significantly without this step. Therefore, we decided not to use this filtering in our restoration process.

SYMMETRY AXIS DETECTION

As seen in section "Objective image model", our image model is symmetrical about a central axis. The restoration method must accurately detect the symmetry axis to restore the damaged side by taking advantage of the other side.
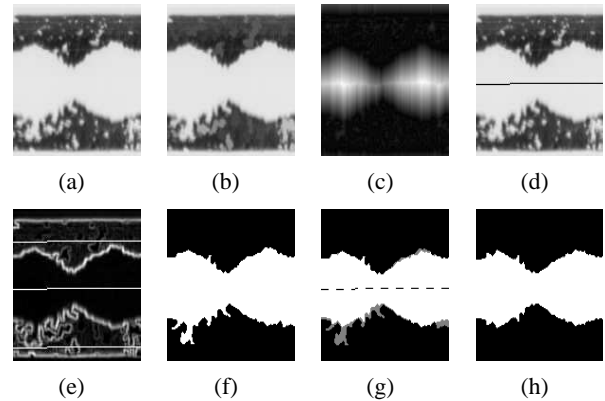

Fig. 2. *Steps of the method proposed in (Brun et al., 2007). (a) Original image. (b) Opening by reconstruction. (c) Quasi-distance. (d) Symmetry axis. (e) Gradient image with markers. (f) Watershed segmentation. (g) Symmetry enforcement and edges correction. (h) Final image*

Yuan and Tang (2005) classified the methods of symmetry axis detection into two main categories. The methods of the first category are based on edges and shapes of objects. The other category mainly uses region-based techniques. In our case, we exclude the first category methods because they require a preliminary segmentation. Some of the methods of the second category require a long computation time, the others are not accurate enough for our task. Moreover, our problem is different from the ones addressed in these papers since we know approximately the position of the symmetry axis and we want to determine it accurately. Therefore, the existing algorithms are not suitable to our task.

To determine the symmetry axis, we used in our previous method the quasi-distance (Beucher, 2007). Unfortunately, in some images of our new database this method gives wrong results, especially when dark pixels are located in the bright region. Moreover, although efficient algorithms to compute the quasi-distance have been developed (Enficiaud, 2007), it still requires a computation time which is too long for our application. Therefore, we propose a new method for the symmetry axis detection.

Since the slope of the symmetry axis is very small, we compute the symmetry point of each parallel line to the y-axis. The linear regression of all these symmetry points gives the symmetry axis. Let L_x denote a parallel line to the y-axis (Fig. 3a). For each pixel p of the line L_x, we compute the mean of the absolute differences between the gray levels of each pair of pixels in L_x symmetrical about p (Fig. 3b). We call the obtained value the *folding* of the line L_x at the pixel p.
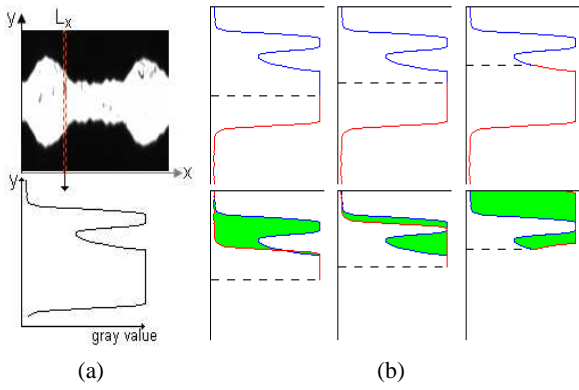
(a)   (b)

Fig. 3. *Folding procedure. (a) Profile of gray levels of a line. (b) Folding of the line at three different pixels: the folding value corresponds to the mean of the filled region.*
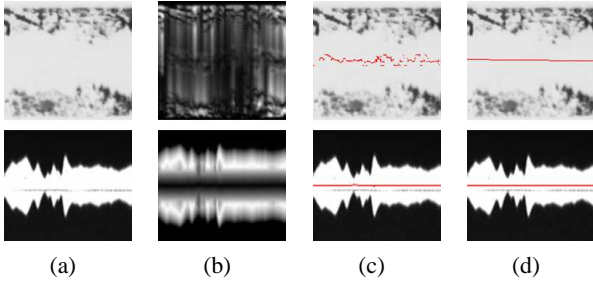

(a)   (b)   (c)   (d)

Fig. 4. *Steps of symmetry axis computation. (a) Original images. (b) Folded images. (c) Crest points. (d) Symmetry axis.*

(a)   (b)   (c)
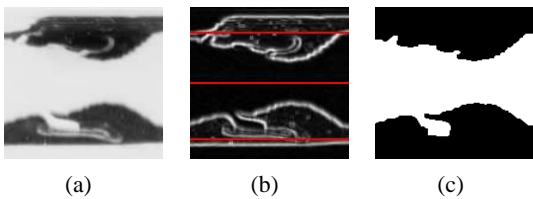
Fig. 5. *Watershed segmentation. (a) Original image. (b) Gradient image with markers. (c) Resulting segmentation.*
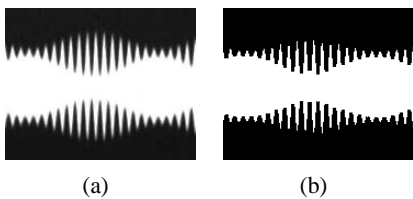

(a)   (b)

Fig. 6. *(a) Image with high frequencies. (b) Watershed segmentation.*

Computing this value for each pixel in each line of an image f gives the *folded image F* :

$$F(x,y) = \frac{\sum\limits_{k=1}^{n_y} |f(x,y+k) - f(x,y-k)|}{n_y} , \qquad (1)$$

where n_y is the number of the computed differences when folding the line L_x at the pixel $p(x,y)$, *i.e.*, the number of the pixels in L_x which are symmetrical about $p(x,y)$.

In contrast to the quasi-distance, the influence of dark pixels located in the bright regions on the folded image is neglected because the differences are averaged. Moreover, the nature of the folding algorithm makes it faster than the quasi-distance one.

To determine the symmetry axis we compute a linear regression of the crest points of the folded image (Fig. 4). Note that the crest points are sought in a small interval around the center of the image. Its length is empirically set to one tenth of the image height. Therefore, the folding image is only computed for this interval of interest and not for the entire image. Moreover, if the value of a crest point (in the folded image) is not small enough, this implies that the two sides are not symmetrical. Therefore, for the computation of the linear regression, we only consider the crest points which have a value less or equal to a certain threshold empirically set.

IMAGE SEGMENTATION

The most important step in our restoration method is the image segmentation. It will determine the quality of the final results. The aim of this step is to segment the image in such a way that the frontiers of the segmentation correspond to the contours of the soundtrack. We want to obtain three distinct parts : the light central region and the two dark lateral regions.

The segmentation cannot be performed with a simple thresholding because of the dark pixels located in the light region and vice versa. Moreover, the value of the threshold is not easy to determine automatically. Standard morphological operators might fill the valleys or erode the peaks. In our previous method we used a morphological watershed from markers on the gradient image. We computed the morphological gradient of the image and an appropriate set of markers (as described in section 3). Fig. 5 shows a soundtrack, its morphological gradient with markers and the associated segmentation.

A well-known drawback of the watershed approach is the inaccurate detection of thin structures. Because of this drawback, the segmentation of high

frequencies images with this approach is not accurate enough. Fig. 6 illustrates this problem.

In contrast to the watershed segmentation, the region growing approach (Adams and Bischof, 1994), is known to preserve thin structures. We have chosen, as in the watershed segmentation, the symmetry axis as marker (or seed) for the central region. But in order to reduce the number of parameters, we have chosen the two lines which have the minimum sum of gray values as side markers. We use a standard region growing algorithm as described in (Adams and Bischof, 1994).

Fig. 7 shows the result of the region growing algorithm on two different images. We notice that this segmentation allows a better localization of high frequency signals (Fig. 7a). However, the resulting segmentation is not always a causal curve and it might contain some unwanted thin structures (Figs. 7b and 7c).

As it is very important to correctly segment high frequency images, we have chosen to use the region growing algorithm instead of the watershed one.

SYMMETRY ENFORCEMENT AND EDGE CORRECTION

The aim of this step is to enforce the symmetry of the segmented image and to obtain causal edges. In order to perform this task, we proceeded in our previous method as follows: in each perpendicular line to the symmetry axis, if a black pixel is separated by a distance d from the symmetry axis, all pixels on the same line, which are separated from the symmetry axis by a distance larger than d become black. This procedure gives generally good results (Fig. 2g). Unfortunately, in some images of our new database, this procedure does not make it possible to detect the damaged side of the soundtrack. Therefore, a more robust method had to be developed.

It is obvious that the total variation (*i.e.*, the integral of the Euclidean norm of the gradient) of the correct side of the soundtrack is smaller than the total variation of its damaged side (see Osher and Rudin, 1990). For this reason, we decided to compute, for each line in the soundtrack, the total variation of its both sides. This is done as follows: Let L_x denote a parallel line to the y-axis (Fig. 8a). We are interested in the total variation in the two lateral sides of L_x which have been segmented as black (Fig. 8b). We compute the gradient of L_x on these two sides (Fig. 8c). The integral of the absolute value of the gradient gives the total variation of each side. By proceeding so, we obtain for each line L_x, two values which characterize the degradation level of each side (Fig. 8e). Therefore, we favor in the symmetry enforcement, the side for which this value is the smallest (Fig. 8f).
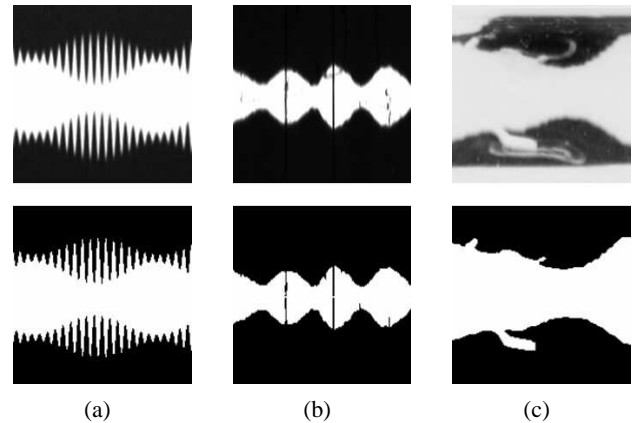

(a) (b) (c)

Fig. 7. *Region growing segmentation examples. (a) Image containing high frequencies. (b) Image containing unwanted thin structures. (c) Image with non-causal edges.*
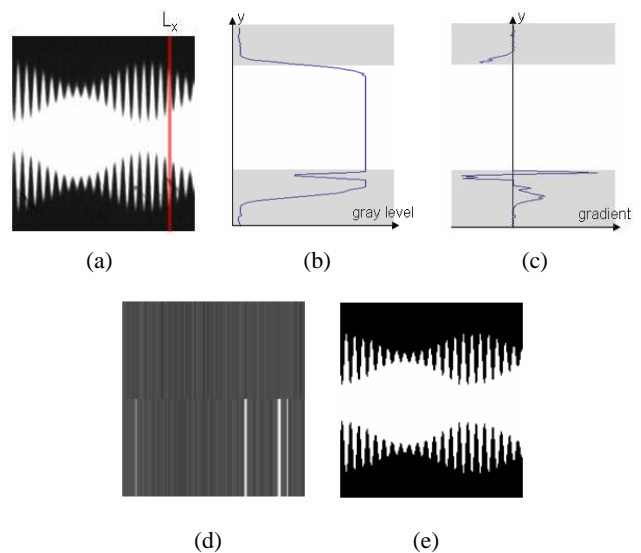

(a) (b) (c)

(d) (e)

Fig. 8. *Computation of the total variation of each side. (a) Line profile. (b) After Segmentation. (c) Gradient of each lateral side. (d) Total variation for the two sides of each line (the brighter the side, the higher its total variation). (e) Resulting image.*

CONTOURS SMOOTHING

Although the obtained binary image is visually satisfactory, its corresponding audio signal contains a background noise which is more audible than in the audio signal of the original image. This is due to the fact that the binarization process restricts the number of possible values of the audio signal amplitude to the height of the image (whereas the number of these values in the original image is *the number of gray levels (256) × the height of the image*). Thus,

the binarization induces a quantization of the audio signal which causes the introduction of audible high frequencies. This phenomena is particularly audible in signals with low frequencies.

Fig. 9 illustrates this fact. It shows a part of a soundtrack and the spectrogram corresponding to its audio signal (Fig. 9a), the corresponding binary image and the associated spectrogram (Fig. 9b). Note that in the case of the binary image, the spectrogram contains parasitic frequencies.

An intuitive idea to undo this effect is to smooth the contours of the image. This is done by replacing the contours and their neighbors in the final image by a weighted sum of the original image and the binary image. The closer the pixel is to the contour, the more the weighting function favors the original image. Only the correct side of the soundtrack is used in this process. More precisely, let L_x denote a parallel line to the y-axis in the original image, B_x the corresponding line in the binary image and y_0 the y-coordinate of the contour in the undamaged side of L_x. The final result of the n neighboring pixels of the contour is done by R_x such as:

$$R_x(y) = \lambda(y) \cdot B_x(y) + (1 - \lambda(y)) \cdot L_x(y), \quad (2)$$

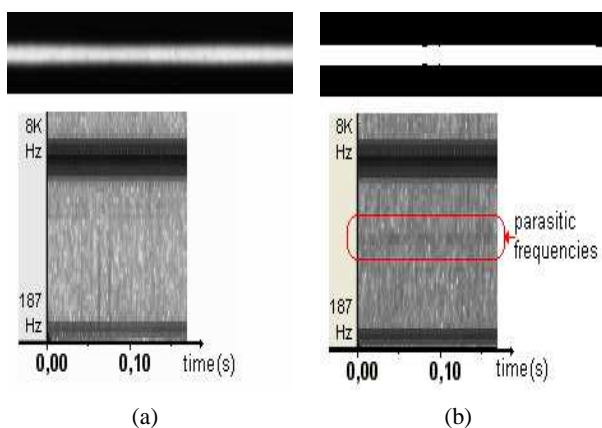where $\lambda(y) = |y_0 - y| / n$ and n empirically set to 6.

Fig. 9. *Binarization effect. (a) Original image and the associated spectrogram. (b) Binary image and the associated spectrogram.*

RESULTS

Figs. 10 and 11 show two restoration examples. In most cases, when only one side is degraded, it is restored thanks to the other side. Note that, in order to have a good visual assessment, we display in addition to the final image, the edges of the binary image on a brighter version of the original image.

Audio samples corresponding to optical soundtracks before and after restoration are available at the project website (http://www.riam-resonances.org/) in the "Audio samples" section.

An important rule that was adopted during the design of the proposed restoration method is to have as few parameters as possible. There are currently only three parameters that need to be chosen by the user: the length of the interval in which the crest points are sought, the gray level threshold from which the crest points are considered and the number of neighboring pixels to take into account in the smoothing process. Their values were empirically chosen and kept constant for all the images.

For processing 100 images having a resolution of 640×512, our restoration method takes 86 seconds, when using a 2.13 GHz Pentium 4 processor with 2 GB RAM. Those images are processed by our previous method in 352 seconds. So, the new method is at least four times faster than the previous one. The difference in the computation times is mainly due to the computation of the quasi-distance and the gradient.

According to motion picture experts, this method gives results which are at least as good as the previous one. In some cases (high frequency signals, important dark artifacts on the central region) the results are even better. Moreover, the resulting audio signal obtained by the new method does not generate parasitic frequencies due to the binarization process. Finally, in contrast to the restoration methods which are performed at the audio domain, our method makes it possible to remove the noise caused by the degradation of the film while conserving the background noise which is present in the original recording of the soundtrack.

Fig. 10. *Example with a very degraded soundtrack. (a) Initial image. (b) Edges of the binary image with a brighter version of the original image. (c) Restored image.*
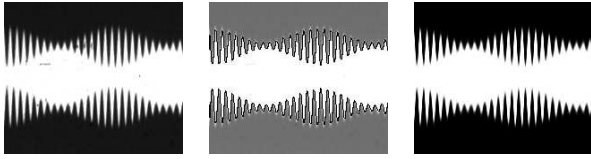
Fig. 11. *Example with high frequency images. (a) Initial image. (b) Edges of the binary image with a brighter version of the original image. (c) Restored image.*

CONCLUSION

We have presented a method for variable area soundtrack restoration. This method is an improvement and completion of a previous one. It gives an accurate detection of the edges in soundtracks with high frequencies and is four times faster than the previous one. We tested this method on several images of our databases and, according to motion picture archivists, the results are very satisfactory. Furthermore, in the resulting audio signal, only the noise caused by degradation is removed.

In addition to the performed audio evaluation, a new acquisition campaign including music and voice is being performed in order to undergo a more complete auditive assessment.

There is currently ongoing work on the treatment of over- and under- exposure of the optical soundtrack and the development of a real-time restoration method for slightly damaged soundtracks. Finally, some defects might not be observed at the image representation of the soundtrack, therefore the detection of anomalies in the corresponding spectrogram needs to be studied.

REFERENCES

Adams R, Bischof L (1994). Seeded region growing. IEEE Trans Pattern Anal 16:641–7.

Beucher S (2007). Numerical residues. Image Vision Comput 25:405–15.

Brun E, Hassaïne A, Besserer B, Decencière E (2007). Restoration of variable area soundtracks. In: Proc 2007 IEEE Int Conf Image Process, San Antonio, Sept 16-19. IV:13–6.

Enficiaud R (2007). Algorithmes multidimensionnels et multispectraux en morphologie mathématique: Approche par méta-programmation. PhD thesis, Ecole des Mines de Paris.

ISO7831 (1986). A-chain frequency response for reproduction of 35 mm photographic sound – reproduction characteristics. International standard.

ISO2969 (1987). B-chain electro-acoustic response of motion-picture control rooms and indoor theatres – specifications and measurements. International standard.

Kuiper A (2005). Detection of dirt blotches on optical soundtracks using digital image processing. In: Proc 15th Int Czech-Slovak Sci Conf Radioelektronika, Brno.

Osher S, Rudin L (1990). Feature-oriented image enhancement using shock filters. SIAM J Numer Anal 27:919–40.

Richter D, Poetsch D, Kuiper A (2003). Localization of faults in multiple double sided variable area code sound tracks on motion picture films using digital image processing. In: Proc 13th Int. Czech-Slovak Sci Conf Radioelektronika, Brno.

Streule P (1999). Digital image based restoration of optical movie soundtracks. MSc thesis, ETH Zurich.

Valenzuela J (2003). Digital audio image restoration: Introducing a new approach to the reproduction and restoration of analog optical soundtracks for motion picture films. In: Proc Int Broadcast Conv 2003, Amsterdam.

Yuan T, Tang X (2005). Efficient local reflectional symmetries detection. In: Proc 2005 IEEE Int Conf Image Process, Genova, Sept 11-14. III:1180–3