Restoration of under/overexposed optical soundtracks

Bernard Besserer¹, Abdelâali Hassaïne² and Etienne Decencière² ¹ Laboratoire Informatique, Image, Interaction Université de La Rochelle, F-17042 LA ROCHELLE, FRANCE

bernard.besserer@univ-lr.fr

² Centre de Morphologie Mathématique

Ecole Nationale Supérieure des Mines de Paris, F-77305 FONTAINEBLEAU

 $\{assaine, decencie\}$ @cmm.ensmp.fr

Abstract

Film restoration using image processing has been an active research field during the last years. However, the restoration of the soundtrack has been mainly 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. Digital restoration of optical soundtracks at the image level is an application field which has not been investigated, though it is scientifically rich. It consists in removing dust particles, concealing large corrupted areas or restoring underexposure or overexposure defects of the soundtrack. Finally, it makes it possible to mix image processing and signal processing approaches.

After introducing the principles of optical soundtrack recording and playback, this contribution focuses on experimental approaches to detect and cancel the effects of under and overexposure. The approach is validated on both simulated alterations and real data digitized using the specific optical soundtrack scanner we have set up for the RESONANCES project.

Introduction

A quite general introduction should be useful, most people do not know how sound is carried for theatrical release prints, the most popular thoughts on this issue would be a separate medium. But over almost 80 years, the sound has been carried along the pictures on the film stock itself, as an optical track, for both analog sound and modern digital sound (Dolby Digital or SDDS¹). We focus in this paper on analog soundtracks which has been used from the thirties until today, and still present on release copies as backup in case the reading of digital data fails (see fig 1).



Figure 1. 35mm film strip showing modern digital soundtracks along the analog VA soundtrack.

Analog optical sound has a narrow dynamic range, as well as a limited frequency response (according to nowadays standards), but early sound was intelligible, often pleasant to listen to, showed incredible interoperability between evolving standards, and the analog soundtrack is somehow robust against impairments. Optical sound recording has indeed an interesting and rich history [4, 5, 6, 8]. Motion pictures have historically employed several types of optical soundtracks, ranging from variable density (VD) to stereophonic variable area (VA) tracks. For many years, the standard industry practice for the 35 mm theatrical release format has been the variable area optical soundtrack, called The standard Academy Optical track and introduced ca. 1938. Between the sprocket holes and the picture, a 1/10" inch (ca. 3mm) is dedicated to the optical soundtrack.

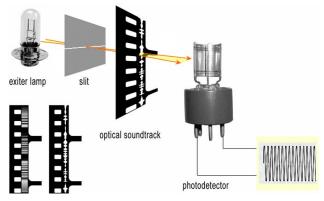


Figure 2. The reproduction process of a VA optical soundtrack.

In general, sound is recorded on the film by exposing this area to a source of light in a so-called "optical recorder". For VD soundtracks, the light intensity of the recorder is modulated and the film density, after processing, goes through varying shades of gray. For VA soundtracks, the width of exposed area is modulated, and the track comprises a portion which is essentially opaque and a portion which is left transparent, the ratio between the two portions being proportional to the instantaneous amplitude of the sound signal being recorded. The reading of the soundtrack consists in the inverted process. A light beam is projected through a slit, then through the film, which continuously streams and therefore modulates the light, while a photoelectric device picks up the amount of light and feeds the amplifier stage, as illustrated fig 2. Note that the same pickup head is able to read VA or VD tracks (in both cases the amount of light varies).

Restoration of optical soundtracks

Unfortunately, the optical soundtrack undergoes the same type of degradations as the image of the film (dust, scratches...etc). Given that they are located close to the film stock edge, soundtracks are sometimes degraded by abrasion in

¹Sony Dynamic Digital Sound

the neighborhood of the perforations or by fungus or mold attacking the film on an important surface. An example of corrupted soundtrack is shown in Fig. 3.



Figure 3. A heavily corrupted soundtrack (fungus or mold).

Classically, sound processing and restoration are performed only after the transformation of the optical information into acoustic electric signal. Impulsive impairments are easy to conceal in the 1-D signal domain, but the presence of large area degradation or repetitive defects on the soundtrack 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.

This has been one of the major reasons which led to the launch in 2005 of the RESONANCES research program. This program mainly aimed at restoration of optical soundtracks in the "image domain". Moreover, there are only few references in the literature on this topic.

In 1999, Streule [12] proposed a soundtrack restoration method using digital image processing tools. He proposes a complete system, going from the soundtrack digitization, up to the generation of the corresponding audio file. Concerning the restoration, Streule only treats defects caused by dust. The proposed technique is mainly based on the soundtrack symmetry.

Richter, Poetsch and Kuiper proposed in [9, 10] a method of impairments localization in multiple double sided variable area soundtracks, but they do not treat the correction of these impairments. This method eliminates low frequencies in Fourier domain, which correspond to small defects in the original image, and after a binarization, the remaining faults are sufficiently large to be easily detected. Spots detection is also used by Kuiper in [7]. The spots being lighter than other parts of the image, a threshold isolates them. A succession of morphological operations is then applied to localize the spots and to remove the isolated pixels. Unfortunately, in most cases, the spots are not lighter than the other parts of the image. For that reason, this method cannot be always used.

J. Valenzuela appears as the inventor of several patents on soundtrack scanning and restoration. He proposes a short description of his technique in [13]. The restoration method is very simple, it is based on median filters and erosions, and it can only deal with small defects.

None of the previous techniques would allow a satisfactory restoration of moderately to severely damaged soundtracks. Moreover, to the extent of our knowledge, nothing has been published on the restoration of incorrectly exposed optical soundtracks.

Optical Soundtrack Alterations: Dust and scratches

Dust spots have a significant impact on variable density tracks; the VD process causes an important background noise, due to film grain and dust spots: every dust particle causes a variation of the intensity. The VA process is much more robust with respect to dust on the dark portions (black over black). This is one of the reason the VD process was replaced by the VA process.

As the quality of the films themselves improved, dealing with noise became critical, especially during the quiet moments of the film. Several so-called **noise gates** were introduced to improve the VA process, biasing the light valve to reduce the amount of light being passed through the track (the width of the transparent area) during moments of silence.

Removing dust, scratches and other defects is one of the aims of the RESONANCES project. Therefore, an advanced image processing method has been developed within this project, in order to remove defects and restore the track symmetry [2]. This method detects the symmetry axis of the soundtrack and segments it using a watershed algorithm, into a light central region and two dark lateral regions. The image thus obtained might not be symmetrical, therefore the final step aims to enforce its symmetry.

This contribution focuses on the correction of over- and underexposed soundtracks. Therefore, we will hereafter assume that we deal with clean and symmetric samples.

Optical Soundtrack Alterations : underexposure and overexposure

Similarly to the images of a movie, the optical soundtrack undergoes several copies, from the masterized soundtrack photographed by the optical recorder to the final print. Therefore, density control is important. Moreover the film stock used and the parameters of the development process also influence film density. The quality control for this production chain was of great importance for variable density soundtracks and hard to manage, and this is another reason for the demise of VD tracks. VA tracks are more tolerant to exposure and development conditions, since the pattern to be reproduced is more or less binary (transparent track, opaque surroundings). But under certain conditions, bad exposure can affect significantly the VA track, due to image spread (or flare) and the S-shaped response of the film : Suppose a small, sharply focused spot of light is exposed on a piece of film ; after processing, the image is likely to be larger than the initial spot of light.

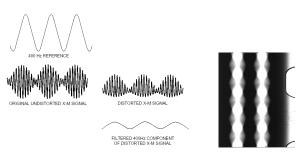


Figure 4. Left : Graphical illustration of the cross-modulation test (lifted from Kodak's technical note "Cross-Modulation Distortion Testing for the Motion Picture Laboratory"). Right : Image grabbed from a real cross-modulation test reel (stereo tracks).

From the forties forward, an industry-standard practice raised, commonly known as the "cross-modulation test", and is still used as a quality assurance routine prior to sound recording and duplication. The test is based on the fact that a perfect sinusoid comprising a high frequency signal (about 10 KHz) modulated at 75% by a low-frequency one (typically 400 Hz) will have an average value of zero (the average light transmission will be constant). In the case of underexposure or overexposure, some of the low-frequency modulation component will be introduced into the average value of the signal and may be detected. This technique is still used and we suggest the eager readers to study further the technical note from Kodak [1] on the cross-modulation test.

At the optical representation of the soundtrack, the effect of under/overexposure changes the shape of the modulation if the frequency is above ca. 5 Khz : a pure sinusoidal wave takes on a sharper, more saw tooth shape, either on the inner side (under-exposure) or the outer side (overexposure), as shown in Fig. 5. On pure frequency signals, the effects of the over-exposure are the same ones as those of the under-exposure (with a phase shift of π). It seems to be very hard and complex, for an arbitrary 1D audio signal, to distinguish between distortion introduced by over-exposure from the distortion introduced by under-exposure, therefore this detection could only be performed in the image domain, or the operator will set the "sign" of the distortion (under or over).

The distortion level is frequency dependant: the image spread does not significantly change low frequency signals (under 1 Khz). The image spread introduces first a desymmetrization of the signal, therefore this distortion generates even harmonics at rather low frequencies (2 to 3 Khz). At higher frequencies, the shape of the signal is altered, introducing moreover odd harmonics.

While listening, voice is mainly affected, especially the sibilants. But such distortion is hardly noticeable for music.

Figure 5. Form left to right : underexposed, correctly exposed, overexposed, real optical soundtrack showing under/over exposure.

The physical phenomenon which causes the over/under exposure can be fairly accurately modeled in the image domain. We have therefore built an exposure simulator : The easiest way to simulate the effect of under/overexposure is to simulate the behavior of the image spread, consequently the simulator deals with the optical representation of the soundtrack as 2D image. The optical spread is simulated by convolution (2D gaussian kernel or 2D squared cardinal sine filter, often used to model the point spread function in astronomcal imagery). The resulting gray levels are matched against a S-shaped (sigmoid) look-up table, roughly simulating the film transfer function. We designed a framework under MATLAB with a suitable user interface, allowing us to calculate the following steps :

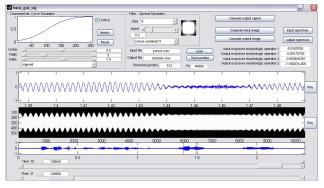


Figure 6. MATLAB user interface of the simulation framework. : we are able to load a WAVE sound, convert it into its optical representation, simulates the image spread and convert the signal back to WAVE.

Restoring underexposed and overexposed optical soundtracks

Only severe under/overexposition can be discerned by looking to the optical representation, and only if some reasonably high frequency tone is present in the signal. The grabbed picture shown in Fig. 5 shows such oversharp peaks. This is an extreme case, and for our project, more gentle distortions should be detected as well. Therefore, we set up two separate paths in our research planning : One approach will deal exclusively with the optical representation of the soundtrack, the second one, described here, will perform the detection step based on the audio signal.

We are looking for a way to automate detection, measurement and reduction of distortion induced by over-exposure (or at least to guide the user of the digital processing chain to perform the correct correction). As stated before, we focus on 1D audio signal for the detection and measurement of the distortion, without reference tone. The correction itself will take place in the 2D image representation of the soundtrack.

We do not believe in a reliable indicator able to characterize (in an absolute manner) how much a sound (of any kind) has been distorted. However a closed loop or an iterative approach may be realistic. In such an approach, a suitable indicator should roughly express the magnitude of the correction. This indicator will be averaged over time and according to this indicator, image correction will take place. This iterative approach should converge to a minimum.

Measuring the distortion in the image domain

The detection of the under/overexposed soundtracks with image processing operators seems to be a promising strategy. Mathematical Morphology [11] offers operators which are well adapted for dealing with this sort of geometrical problem. It is interesting to note that the effect of the overexposition of a soundtrack seems to be similar to the effect of the application of a morphological dilation with a certain structuring element. According to mathematical morphology theory, if this hypothesis is true then the soundtrack should be invariant to the application of a morphological opening with the same structuring element. The structuring element is a priori unknown. Given the physical process that causes overexposure, it can be safely supposed that it is a disk. Several sizes (limited by the discrete nature of the scanned soundtrack) should then be tested. However, we can anticipate that the presence of noise (film grain, dust, etc.) might interfere in the verification of the hypothesis. Therefore, we have pre-processed the images of the soundtracks using the method described in [2] in order to binarize them and suppress the noise.

The application of a series of openings with structuring elements of increasing sizes allows us to check the invariance conjecture. Note that in the case of soundtracks only containing low frequency signals the invariance is always observed, given that such tracks do not contain thin structures, whose shape is subject to variations when overexposed. If a different behavior exists, it can only be observed in the case of high frequency signals (see Fig. 5). In such cases, we have indeed observed a near-invariance through a morphological opening, which tends to confirm our hypothesis. Note that the detection of underexposed soundtracks could be done in exactly the same way, by previously inverting the binary image of the soundtrack.

However, extra work needs to be done in order to be able to validate this detection method. For instance, the knowledge of the frequencies present in the soundtrack should be used to trigger the use of this detection method.

Once overexposure has been diagnosed, a correction is necessary. This could also be done in the image domain using mathematical morphology. In fact, we have seen that the detection of the overexposure also produces the size of the structuring element used in the dilation which models the overexposure. It will be seen in section "Correction by mathematical morphology" how this can be done.

Measuring the distortion in the audio domain

We need an indicator computed from the 1-D audio signals, able to determine whether or not a sound sample has been distorted (due to incorrect exposure). Usually, distortion is expressed in relation to a reference signal. So we first looked for pitch detection to automatically extract a reference, but we rapidly noticed that this will be impossible, especially for music.

After discarding other methods, we propose in this contribution two possible approaches :

Spectrum based indicator As an incorrect exposure introduces more harmonics for high frequencies, one of the considered approaches was to compute the Center Of Gravity (COG) of Spectrum (in a sense, the "mean" frequency, and this method is used for pitch detection and for audio restoration [3]). It is calculated piecewise for different frequency ranges by (vis an array indexed from 1 to N) :

$$COG(v) = \begin{cases} 0 & \text{if } \sum_{n=1}^{N} v(n) = 0\\ \sum_{n=1}^{N} v(n) \times n\\ \sum_{n=1}^{N} v(n) & \text{else.} \end{cases}$$

We compute the COG for different ranges, increasing the amount of high frequencies in the calculation. So we expect seeing the curves drifting apart if distortion is present. The COG-based indicator, which intends to reflect the importance of under/overexposure, is computed by summing the distance between all possible couples of the K COG tracks, as :

$$indicator(t, cog_K(t)) = \sum_{k=1}^{K} \left(\sum_{l=k+1}^{K} |cog(t,k) - cog(t,l)| \right)$$

where $cog_K(t)$ are the *K* COGs that have been computed at time *t* of the signal.

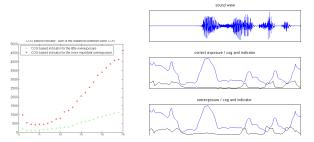


Figure 7. Left : COG-based indicator plotted over time for the frequencysweep input. As expected, the indicator rises as frequency increase. Right : COG plot (blue) and COG-based indicator (black) for a real sound sample. Even if the variation is small, it is effective over the complete sample.

As COG is one of many known techniques for pitch detection, the ensued indicator somehow follows the pitch of the sound sample. To be used as feedback value in our closed-loop approach, a low-pass filtering / averaging has to be applied to this value. This is not a problem, as under/overexposure effect is constant over a long period (a complete reel, or at least over a shoot, if there are several parts spliced together on the reel).

Note that noise disturbs this method, especially impulsive noise which creates high frequencies, thus rise the COG. Fortuitously, impulsive noise is easy to remove in the image domain (dustbusting).

- Harmonic distortion based indicator Since the distortion adds principally even harmonics on high frequencies, this indicator should reflect the harmonic distortion (even harmonics) for supposed fundamental frequencies, if present. Total Harmonic Distortion (THD) is often used to characterize audio equipment (non-linearities, non-symmetric stages, clipping). The analogy to our problem (desymmetrization, clipping) is great enough to undergo a trial. But THD is measured by feeding the equipment with a fixed and known signal. Since our signal is recorded without any reference, our approach consists in the following steps : the input signal is filtered with a filter bank. Each filter selects one supposed fundamental frequency. Then, for each presumed fundamental frequency, we compute the energy of the odd and even harmonics relative to each supposed fundamental frequency, up to $f_e/2$, using two comb filters for this selection. Instead of computing a ratio between harmonics energy and the energy of the presumed fundamental frequency, we multiply the values. Therefore:
 - if the presumed fundamental frequency is of low energy, this cancels out the harmonics energy, probably not induced by distortion of this fundamental,
 - if the energy of all the harmonics for a given *supposed* fundamental frequency is very low, the resulting value is also low; there is probably no distortion.

Cancels out high degree of harmonics induced background noise during MOS (Moment Of Silence).

• the value is expressive if there is some energy at the presumed fundamental frequency AND there is some energy at harmonic frequencies. This value is also weighted by the amount of selected harmonics (the number of teeth of the comb filter).

These supposed fundamental frequencies have been arbitrary chosen, keeping in mind a future fast IIR implementation. Our set contains following frequencies (in Hz) : 192 240 480 750 1200 1600 2000 3000 4000 4800 6000. Filter design for both bandpass filters and comb filters has been done thanks to MATLAB's filter design tool.

We plot these "harmonic distortion" values against time for several signals (frequency sweep, voiced signal, music) before and after alteration by our simulator, and similarly to the COG based indicator, we combined the results in order to find an indicator which reflects the distortion introduced by a faulty exposure.

Distortion Indicator (DI)

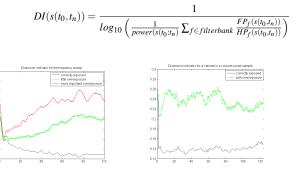


Figure 8. Left : HD-indicator for the frequency sweep test signal (black : correct exposure, green : light overexposure, red : strong overexposure). Right : HD-indicator for music instrument (clarinet) sample (black : correct exposure, green : light overexposure).

Correction of the 2D optical representation of the soundtrack

A very simple correction has been set up to experiment our "closed loop" solution (measurement from the 1D audio output steers the 2D image correction). To do this, the images are grabbed with a great dynamic range (12bits/pixel) together with a fine tuning of lightning power and camera integration time. Consequently, we are able to change the intensity levels of the image pixels over a great range. The computed distortion indicator acts as feedback and helps to set up a simple transfer function (nonlinear look-up table). The transfer function is then applied on the input images and a new iteration is computed.

The advantage of such a method is that even if it could take a long time to determine the best parameters for the transfer function, once it is done, the soundtrack can be corrected and converted to audio signal in real time. The major drawback of such an intensity & contrast adjustment (even non-linear) is that the final result is a binary image, and this causes aliasing effects on the picture. Since the spatial resolution is limited, some high frequencies are introduced and perturb our measures. This problem might be avoided by using super-resolution.

Correction by mathematical morphology

Considering real data, the visual examination of the digitized images advise us that a simple correction based on a transfer function should not be sufficient. We have supposed in section "Measuring the distortion in the image domain" that overexposure can be modeled as a morphological dilation, and we have explained how to validate this hypothesis and compute the size of the corresponding structuring element. If this hypothesis is true, then the theory of mathematical morphology tells us that some information might have been lost in the process, and that a good candidate for the restoration is obtained with a morphological erosion using the same structuring element. Underexposed soundtracks would be restored analogously, by using a dilation.

Therefore, we have applied an erosion to the underexposed soundtrack and generated the corresponding audio signal. We have noticed that the spectrogram of this audio signal contains less high frequencies than the spectrogram of the original audio signal (see fig 9). This result confirms the fact that an erosion filters the high frequencies which have been caused by the underexposure.

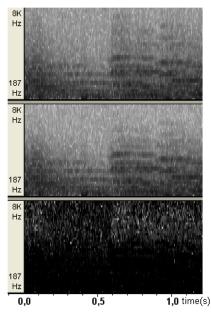


Figure 9. From top to bottom : spectrogram of the underexposed soundtrack, spectrogram of the eroded soundtrack, normalized difference between the spectrogram of the eroded soundtrack and the underexposed one.

Results and Assessment

For the described research, a specific scanner has been built around a reformed sepmag player² in order to start a large scale acquisition and restoration campaign and to validate the method for a very broad set of problems. The optical sound is oversampled at 48KHz by a line scan camera, fitted with a reverse-mount Scheider-Kreuznach macro lens. The film stock is illuminated by a fiber optic line light guide. The size of the resulting image is

²sepmag player : A device able to read sound recorded as separate magnetic tapes (magnetic coated 35mm or 16mm film stock).

 48000×512 pixels for a second of sound. The rather poor line resolution is somehow compensated by a 10 to 12 bits/pixels dynamics to capture the luminance levels along the transition edges of the VA modulation.

Assessment is a difficult task, since we often don't hold any unaltered counterpart to compare with. At current state of the project, validation has been performed on simulated and real data, in a rather subjective way : in fact, it is easier at this stage to do a visual assessment of the restored images, rather than listening to the converted sound. An absolute improvement is hard to perceive while listening to a real altered sound sample, especially for exposure correction. Therefore, a high fidelity audio assessment system must be developed.

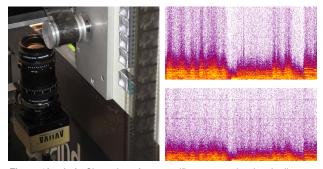


Figure 10. Left: Close shot of our specific scanner, showing the line scan camera and macro lens. Top Right: Spectrogram of an underexposed sound-track, digitized by our scanner and converted to sound. Bottom Right: Spectrogram of the same sample after exposure correction. Notice the noise level for real soundtracks (here no dust removal was performed).

In a near future, we will set up a listening test at a postprocessing auditorium. We intend to follow a blindfold methodology based on selected test tracks. Beside test samples, untouched and restored sound samples from our vintage stock, particular samples will be generated to assess exposure correction: Test tracks (tones, voice and music) will be optically recorded, then copied with a correct exposure, a noticeable underexposure and a noticeable overexposure. Each copy will be converted to sound on our scanner and restored version of the over-/underexposed tracks will be computed and added to the samples.

Conclusion and forthcoming work

Even though image-processing approaches have already been initiated, the goals of the RESONANCES project and the complementarity of the partners will allow us to investigate more deeply the restoration of optical soundtracks, both in the 2Dimage domain and in the 1D audio signal domain. Moreover, the processing of distortion induced by incorrect exposure of photographic soundtracks opens up an unmarked application field, by coupling 1D signal processing and image processing.

Further improvements are especially expected in the 2Dimage domain processing: over/underexposure detection and image correction using morphological operators.

Acknowledgements

This work has been undertaken within the *RESONANCES* project (http://www.riam-resonances.org) and has been made possible thanks to the financial help of the French *Agence Nationale de la Recherche*,

through its *RIAM* program. The film material, as well as the expertise on motion picture optical soundtracks, has been provided by N. Ricordel from the *CNC* - *Archives Françaises du Film* and by C. Comte from GTC-Eclair Group.

References

- Cross-modulation distortion testing for the motion picture laboratory. Technical report, Eastman Kodak Company, 2001. Available at http://www.kodak.com/US/plugins/acrobat/en/motion/ support/h44/h44.pdf.
- [2] E. Brun, A. Hassaïne, B. Besserer, and E. Decencière. Restoration of variable area soundtracks. In *Proc. of the* 2007 International Conference on Image Processing (ICIP-2007), Texas, USA, 2007.
- [3] Andrzej Czyzewski and Przemyslaw Maziewski. Some techniques for wow effect reduction. In *Challenges in Restoration for Media Production*, number TP-L1.8. IEEE International Conference on Image Processing, September 2007. reviewed in April.
- [4] W. Kellog Edward. History of sound motion pictures. In *Journal of the SMPTE*, June, July, August 1955.
- [5] J. G. Frayne and H. Wolfe. *Sound Recording*. John Wiley and Sons, New York, 1949.
- [6] Erpi ClassRoom Films Inc. Sound recording and reproduction (sound on film). An instructional sound film, 1943. Available at http://www.archive.org/details/SoundRec1943.
- [7] A. Kuiper. Detection of dirt blotches on optical sound tracks using digital image processing. In 15th International Czech - Slovak Scientific Conference Radioelektronika, 2005.
- [8] James Monaco. *How To Read a Film*. OXFORD UNIVER-SITY PRESS, Oxford New York, third edition, 2000. available at http://www.readfilm.com/HTRBook/HTR2.pdf.
- [9] Poetsch, Richter, and Kurreck. Restoration of optical variable density sound tracks on motion picture films by digital image processing. In *Proc. of OPTIM 2000*, Brasov, 2000.
- [10] D. Richter, D. Poetsch, and A. Kuiper. Localization of faults in multiple double sided variable area code sound tracks on motion picture films using digital image processing. In 13th International Czech - Slovak Scientific Conference Radioelektronika, 2003.
- [11] Jean Serra. Image Analysis and Mathematical Morphology - Volume I. Academic Press, 1982.
- [12] Patrick Streule. Digital image based restoration of optical movie sound track. Master's thesis, Electronics Labs, Swiss Federal Institute of Technology, Zurich, March 1999.
- [13] J. Valenzuela. Digital audio image restoration : Introducing a new approach to the reproduction and restoration of analog optical sountracks for motion picture films. In *IBC 2003 Conference Papers*. Technicolor Creative Services., 2003.