Zurich University
of Applied Sciences

School of Engineering

MASTER OF SCIENCE
IN ENGINEERING

Master's Thesis
Information and Communication Technologies

Restoration of Delaminated and Cracked Recordings Using Synthetic Depth of Field

A thesis submitted in partial satisfaction of the requirements for the degree of Master of Science in Engineering

Supported by the Lawrence Berkeley National Laboratory
Under the supervision of
Carl Haber, PhD and Earl Cornell, PhD

Written by
Simon Marti

Under the direction of
Prof. Hans Dermot Doran

Berkeley CA USA, 29th July 2015

Author Simon Marti
Institute of Embedded Systems
Zurich University of Applied Sciences
Technikumstrasse 22
8401 Winterthur
Switzerland
martisim@students.zhaw.ch

Main Tutor Prof. Hans Dermot Doran
Institute of Embedded Systems
Zurich University of Applied Sciences
Technikumstrasse 22
8401 Winterthur
Switzerland
hans.doran@zhaw.ch

Second Tutor Carl Haber, PhD
Lawrence Berkeley National Laboratory
Physics Division
1 Cyclotron Road
Berkeley, CA 94720
USA
chhaber@lbl.gov

Edition Draft, February 2015

Git Revision f8438e784fc736764d6ca5348bf3cb3142ca7764

Date 02/27/2015

Abstract

Historical recording media contain valuable information about their time. Restoring the sound information to preserve the content of these media is a high priority for musea and archives around the world.

In 2003 a team a the Lawrence Berkeley National Laboratory under the lead of Carl Haber started to develop an optical scanner for mechanical sound carriers. The system named IRENE allows contactless digitization of cylindrical media and disc recordings. One type of media that remains problematic for the scanning process are delaminating lacquer discs. These discs are made of an aluminum base plate coated with a thin layer of nitrocellulose lacquer, that tends to lift from the base over time. This delamination process creates cracks in the disc surface, which makes it hard to maintain focus during the scan.

Currently the focus is maintained by controlling the distance between disc and camera. This thesis investigates a new method called focus stacking. In this approach the system determines the minimum and maximum height of the disc and then scans over this range with a constant spacing. In the resulting stack of images, every image contains parts of the disc in good focus. Subsequently the stack can be fused into a single image with synthetic depth of field.

The feasibility of using focus stacking is proven through the first implementation. The method provides a higher level of detail, especially in areas around cracks in the disc surface and could improve the scanning process for delaminating lacquer disc in the future.

Acknowledgements

First of all I would like to Carl Haber for giving me the opportunity to write my master thesis on the IRENE project. guiding me in the right direction throughout my work on the thesis It was an honor and a pleasure to work at such an outstanding research insitution as the Lawrence Berkeley National Laboratory

Further thanks go to Earl Cornell for providing ideas and discussing the challanges of the project with me. Answering all my questions and and answering all my questions to the IRENE system.

I would also like to thank Hans Dermot Doran for encouraging me to write my thesis abroad and his help to get the position at the Berkeley Lab.

My parents Erika and Felix Marti for supporting my throughout all my studies Viviane Ludescher for her patience and loving support, even though i decided to write my thesis on the other side of the planet

Erklärung betreffend das selbständige Verfassen einer Masterthesis an der School of Engineering

Mit der Abgabe dieser Masterthesis versichert der/die Studierende, dass er/sie die Arbeit selbständig und ohne fremde Hilfe verfasst hat

Der/die unterzeichnende Studierende erklärt, dass alle zitierten Quellen (auch Internetseiten) im Text oder Anhang korrekt nachgewiesen sind, d.h. dass die Masterthesis keine Plagiate enthält, also keine Teile, die teilweise oder vollständig aus einem fremden Text oder einer fremden Arbeit unter Vorgabe der eigenen Urheberschaft bzw. ohne Quellenangabe übernommen worden sind.

Bei Verfehlungen aller Art treten der Paragraph 39 (Unredlichkeit und Verfahren bei Unredlichkeit) der ZHAW Rahmenprüfungsordnung sowie die Bestimmungen der Disziplinarmassnahmen der Hochschulordnung in Kraft.

Ort, Datum:

……

Unterschriften:

……...

……

……

Das Original dieses Formulars ist bei der ZHAW-Version der abgegebenen Masterthesis zu Beginn der Dokumentation mit Original-Unterschriften und -Datum (keine Kopie) einzufügen.

Contents

List of Figures

List of Tables

Listings

Part I.

Introduction and Background

1. Introduction

1.1. Context

The first known sound recordings were made by the French inventor Édouard-Léon Scott de Martinville in 1857, over 158 years ago [1]. Since these days enormous collections of music, radio broadcasts and other historical documents have been preserved on all kinds of media. Until the advent of magnetic tape recordings in the 1940s mainly mechanical carriers were used to store sound. Many of these are archived in libraries and museums due to their historical value.

The mechanical carries tend to be susceptible to damage, be it through environmental influences or disintegration of the material itself. A great deal of these recordings can not be played back anymore, because either the playback apparatus is not longer available or the mechanical stress of the playback process would destroy the recording. This creates a demand for methods to extract the sound information from these records without further damage.

In 2001 Stefano Cavaglieri from the Swiss National Sound Archives and Prof. Ottar Johnsen and Prof. Frédéric Bapst from the School of Engineering and Architecture of Fribourg, Switzerland presented the VisualAudio method to retrieve the sound from disc recordings using photography [2]. For this purpose a picture of the whole disc is taken with an analog camera. The film is then scanned and the audio can be extracted from the image through image processing and signal processing methods.

Starting 2003 a team under the lead of Carl Haber at the Lawrence Berkeley National Laboratory (LBNL) developed methods to extract the sound from a variety of different mechanical carriers using an optical scanner in the IRENE project [1]. In contrast to the VisualAudio method, in IRENE the sound carrier is imaged directly by a digital scanner. Two separate processes for scanning of cylindrical media or discs have been developed. In both cases the restoration process takes place in two steps, first scanning of the medium and secondly extraction of the audio information. For disc recordings the steps are represented by the IRENE 2D scanning method based on a line scanning camera and the RENE software for audio extraction. Cylinders are scanned with the IRENE 3D system using a 3D probe and the audio extraction is then made with the PRISM software. Through the developed methods several valuable historical recordings could be restored, including Scott de Martinville's which were never played before.

Despite its impressive capabilities, there are still classes of records that present a challenge for the current IRENE system. One of them are lacquer or acetate discs that suffer from delamination [3]. These discs are build up of an aluminum or glass carrier plate coated with a thin layer of nitrocellulose lacquer, which gives them their name. Many of these lacquer discs suffer from aging effects which caused the coating to delaminate. Furthermore the material is not designed for repeated playback.

In contrast to shellac and vinyl disc which are pressed, lacquer discs are recorded directly by cutting the groove into the lacquer. That made them suitable for use in studio and field recordings and to archive radio broadcastings. The trade off for the possibilty of directly engraving

[1] The acronym IRENE was inspired by the American folk standard "Goodnight, Irene" and stands for Image, Reconstruct, Erase Noise, Etc.

a groove is that these disc suffer from wear through the playback stylus very quickly. Lacquer disc were in common use between the late 1930s and the late 1950s. As they were mainly used for direct recordings only a single copy exists of many of them.

1.2. Problem

Lacquer disc often suffer from several aging effects which causes the nitrocellulose coating to separate from the aluminum or glass base. The lacquer layer tends to give of softeners and dry out over time. This results in delamination of the coating from the base material that often occurs in the form of cracks or flaking of the lacquer and renders the disc unplayable [4]. If the records are not stored under constant temperature, the delamination effect can be amplified by the fact that the thermal expansion rate of nitrocellulose is about 4 to 6 times higher than the one of aluminum [5]. The thermal stress can cause the surface to stretch and tear. The delamination usually starts of with small cracks in the coating and can progress until the lacquer flakes of the base in small pieces [3]. Through this aging process the disc surface becomes increasingly uneven which makes it hard to maintain focus during the scan.

The depth of field of the optics currently used in the IRENE system is 28 µm [6]. Disc recordings usually have a warpage that is way higher than this. Therefore IRENE uses active auto focusing based an a laser displacement sensor. This system was designed to keep focus over slightly warped disc but it is unable to adjust the depth of field fast enough if a large number of discontinuities occur.

In 2013 Silvan Fischer developed a method to maintain focus over radial cracks that occur in shellac discs [7]. The method is targeted at cases where the discs is broken into a small number of relatively large pieces and is not well suited for the large number of small cracks that delaminating lacquers show (see section 1.3 for a more detailed description). Therefore an alternative method to generate focused scans of disc with a large number of small cracks has to be found.

1.3. Previous Work

In the 12 years since the IRENE project started a considerable amount of people, many of them students, have contributed to the IRENE project. For the sake of brevity this section only mentions the works that are most important for this project.

The fundamental principles used in the IRENE project to extract the audio signals from the image of a groove are described in the paper *Reconstruction of Mechanically Recorded Sound by Image Processing* by V. Fadeyev and C. Haber in 2003 [8]. As the playback signal of disc recording is proportional to the transverse velocity of the playback stylus this velocity has to be determined. Therefore the lateral displacement of a virtual stylus is derived from the shape of the groove. In the next step the stylus velocity is extracted from the displacement by numerical differentiation. The paper furthermore contains a description of various noise sources, the important physical characteristics of disc recordings as well as a proof of concept.

In 2013 Jeremy Singy addressed the problem of tracking grooves on damaged recordings during his master's thesis *Cracked records with 3D/IRENE* [9]. To be able to extract continuous audio the grooves interrupted through cracks have to be re-linked before the audio extraction. Singy developed software that assists the user by manually tracking the grooves across cracks. This interactive tracking was then successfully tested on disc recordings as well as cylinders. Further-

Figure 1.1.: Laquer disc showing delamination (image source: [3])

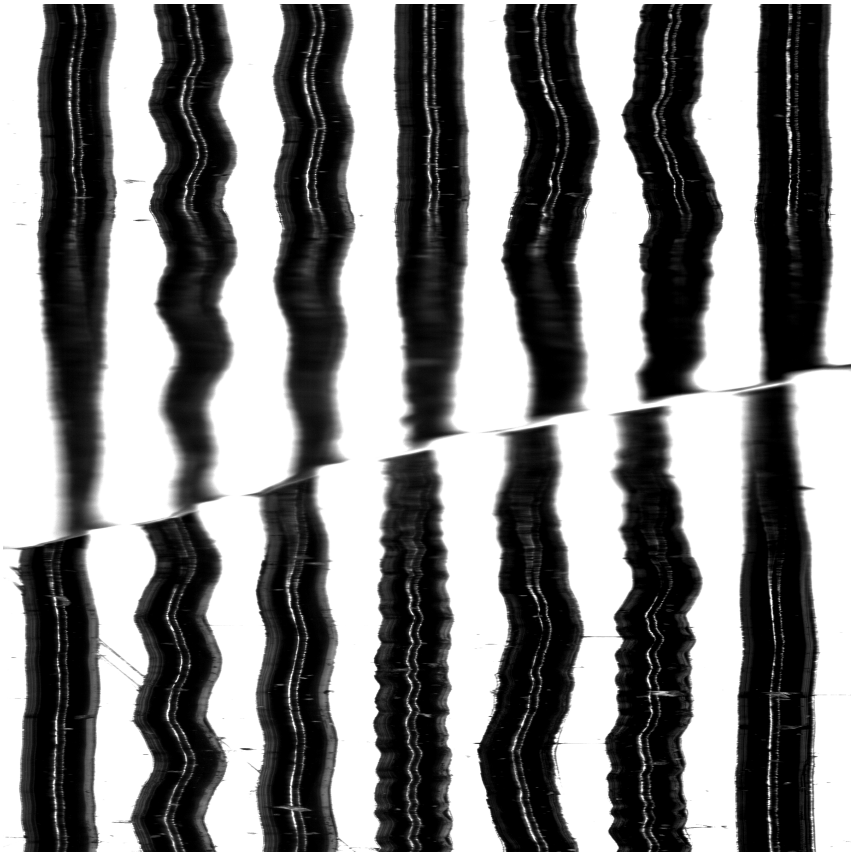


Figure 1.2.: Crack in a section of a scan of a laquer disc

more he investigated the possibility of a fully automatic processing of cracked recordings.

Silvan Fischer developed a method to maintain focus across breaks in shellac discs in his master's thesis *Restoring Cracked Early Recordings* also from 2013 [7]. The method was targeted at cases where the discs is broken into a small number of relatively large pieces and the cracks run in roughly radial direction. The method detects crack in the disc by analyzing the output of the previously mentioned laser displacement sensor. If a crack is detected the blurred region is estimated and the scan is repeated with an adapted camera trajectory. The resulting multiple scans are merged later in the audio extraction software. To track the grooves across the cracks Fischer developed two relinking methods. A fully automated method suitable for disc with a small number of cracks and a semiautomatic one where the user is assisted by the software for relinking the grooves by hand.

1.4. Project Goal

The project aims to investigate the use of synthetic depth of field techniques to handle large focus variation in the 2D scanning of disc recordings. The methods are an interesting research topic, because they could enable the scanning of delaminated and cracked lacquer discs. If suitable methods can be found, the final goal is to extend the current 2D scanning process consisting of IRENE 2D for scanning and RENE for audio extraction with these new methods.

The detailed project goals are divided according to the two main parts of the project. The first goal is to develop an acquisition system that acquires a stack of images with different focal planes. In the second step an algorithm that can fuse the stack of images into one image with synthetic depth of field needs to be designed. Finally if a suitable algorithm can be found, the new method has to be integrated in the current audio extraction process.

1.4.1. Image Acquisition

IRENE 2D Upgrade

The acquisition method developed by Fischer relies on a camera that was replaced in the meantime. The reason for the replacement were issues with the image quality of the camera. To enable the use of Fischer's method with the new camera his code has to be updated to work with the current hardware. Additionally this would allow a comparison of Fischer's acquisition method and the new to be developed one.

Study of Focus and Depth of Field

Before the acquisition system is developed the spacing of the scans has to be determined. Therefore the depth of field of the optical system has to be known. That includes investigations on how offset from the focusing distance affects the scan. A systematic study shall be conducted to determine the depth of field of the current system as well as the influential parameters. If possible the depth of field shall be determined theoretically as well as experimental.

Development of a New Acquisition System

The basic approach for the acquisition system is to take several images of the disc over a range of height offsets, resulting in multiple scans with different focal planes. To determine the necessary

scanning range the height variation of the disc has to be measured in the first step. Afterwards a stack of scans over the range from minimum to maximum height is acquired. The image acquisition system should be designed based on the current IRENE hardware. The output of the image acquisition shall be a raw stack of images. The fusion of images will be done in the RENE software.

1.4.2. Image Fusion

Literature Research

A variety of similar approaches are well established in in macro photography and microscopy under the term focus stacking. Hence literature research for these methods should be done, to gain an overview over the field and eventually identify methods that are suitable for processing the scans.

Development of an Image Fusion Algorithm

An algorithm that creates a fused image with a synthetic depth of field from the stack of subscans needs to be developed. A variety of different approaches should be implemented for evaluation. Finally the most promising method has to be be identified for the integration in the RENE software.

Computational Efficiency of the Image Fusion

Because of the high resolution of the scans the output images have usually a file size between 300 MB and 800 MB [2]. Therefore a computationally efficient fusion algorithm is preferable to keep the processing time in a reasonable range.

Integration of the Image Fusion in RENE

The algorithm shall be integrated in the current RENE software in such a way that the existing audio extraction methods can be applied to the fused image. This way the wide portfolio of existing groove tracking, relinking and audio extraction methods can be used together with the new scanning method.

1.5. Thesis Boundary

The thesis is limited on disk like recordings that can be scanned with the IRENE 2D system. For the development of the new scanning methods only lacquer discs will be targeted, as the scanning process of these discs will probably benefit the most of the new method.

The image acquisition should be based on the current hardware setup of the IRENE system, no additions to the hardware setup will be considered.

[2] The width of one pixel in the image correspond to roughly 0.5 µm on the disc surface, i.e. the resolution of the scan is about 50'000 DPI. One image displays the surface of a 2 mm wide circular ring. Thus the scan of a whole disc result in several 10 of GB of data.

The functionality of the software will be proven by demonstration and evaluation of the output. No dedicated test software will be developed during this project.

1.6. Risk Management

Table 1.1 contains the risk assessment for the project including strategies to mitigate the risks and countermeasure to be applied if necessary.

Image Acquisition

Risk	Probability	Impact	Mitigation	Countermeasures
IRENE 2D upgrade takes a substantial amount of the total project time.	high	The following project steps can not be finished on time.	Asses the current implementation and estimate the duration of the necessary work. Define mandatory and complementary features.	Skip the implementation of complementary features.
A narrow depth of field makes focus stacking impractical.	medium	The proposed method is not applicable. The project goal can not be accomplished.	Determine the depth of field of the current system before starting the work on the acquisition system.	Investigate methods to increase the depth of field of the optical system. Search for alternative methods to scan delaminating lacquer discs.
A large height deviation of typical lacquer discs makes focus stacking impractical.	medium	The proposed method method is not applicable. The project goal can not be accomplished.	Determine the height deviation of typical lacquer disc before starting the work on the acquisition system.	Investigate methods to handle larger height deviations, e.g. determine the surface topography of the disc in advance and adapt the camera trajectory.
The proposed acquisition method is to complicated to be implemented in the given time	low	The following steps can not be finished on time.	Define the minimal feature set for a proof of concept.	Skip the implementation of complementary features.

Image Fusion

Risk	Probability	Impact	Mitigation	Countermeasures
The proposed method does not deliver usable results.	medium	The project goal can not be accomplished.	Undertake a literature research to get an overview over existing methods for focus stacking. Evaluate several method for synthesizing the depth of field.	Provide a detailed problem analysis.
The proposed method is to complicated to be implemented in the given time	low	The thesis can not be finished on time.	Define the minimal feature set for a proof of concept.	Skip the implementation of complementary features.
The method can not be integrated in the current RENE software	low	The image fusion can not be used with the current tracking methods.	Assess the requirements for an integration of the new feature in RENE before starting the development of the image fusion.	Implement a stand alone proof of concept.
The existing tracking methods do not work with the merged images.	low	The existing tracking algorithms have to be adapted or new ones have to be developed.	Asses the requirements for using the existing tracking methods before starting the development of the image fusion.	Investigate alternative tracking methods. Implement a stand alone proof of concept.

Table 1.1.: Project risk assesment

1.7. Data Management and Version Control

Both projects IRENE 2D and RENE are currently each stored in a separate Git repository which are hosted on Bitbucket [3]. At the end of the project the new features should be integrated into these existing repositories. The repositories are available at the following links:

- IRENE 2D: https://bitbucket.org/ireneteam/irene_2d
- RENE: https://bitbucket.org/ireneteam/rene

IRENE 2D contains mainly LabVIEW code stored in .vi-files, a proprietary binary format, which makes branching and merging rather difficult. Therefore all changes on IRENE will be made directly on the main branch. Presumably no one else will be working on the IRENE 2D code during the course of this project, so the risk of conflicting commits is very small.

RENE is written completely in C#, thus the normal Git workflow can be applied without any problems. A feature branch workflow, where every new feature is developed in new branch, is used here. This workflow enables someone else to contribute to RENE during the course of the project, while remaining a simple workflow [10].

All code developed solely for this thesis and the sources of the documentation are stored in separate repository, which can accessed via the following link:

https://bitbucket.org/simonmarti/lacquerdiscsdaq

1.8. Report Structure

This report is divided into five parts, introduction and background, image acquisition, image fusion and tests and conclusion. The first part contains the introduction, including the project goals, a chapter about the historical context of the project as well as description of the current IRENE system. The second part contains the IRENE upgrade, the result of the study to determine the depth of field of the current system and describes the implementation of the new control software for scans with multiple focal planes. The image fusion part gives an overview over related work on the topic of focus stacking and deals with the implementation of the image fusion algorithm. The last part explains the conducted test and their outcome and finishes with conclusions and an outlook on further work.

1.9. Summary

The aim of this thesis is to investigate ways to handle the focus variation when damaged discs are scanned. The search for alternative ways to handle the focus is motivated by the fact, that lacquer disc suffering from delamination can not be scanned with the current IRENE system. An acquisition system that takes several scans of the the disc with different focal planes shall be developed. In a second step the scans will be merged into one single image with a synthetic depth of field.

[3] Bitbucket is a hosting service for Git and Mercurial repositories (https://bitbucket.org/).

2. Historical Background

"The first person to capture sound in the field was the ethnographer Jesse Walter Fewkes who, on March 15, 1890, walked out into a field in Maine to record a Passamaquoddy Indian harvest song. His medium was wax; his recording device of choice (there were no others at the time) was an Edison wax cylinder, very similar to the Singer pedal sewing machine, in which one drives a belt with repeated foot movement. Since that fateful day we have imprinted sound on tin, wire, glass, acetate, and magnetic tape. Each of these media has its own set of problems: impurities in manufacturing, innate poisons in the substances, exposure to air, and improper storage conditions that lead to decay and rot. Ultimately the death knell is sounded for all of these recordings." Mickey Heart, 2001 [11].

2.1. A Brief History of Sound Recording

The first know sound recordings were made by the French inventor and printer Édouard-Léon Scott de Martinville [12]. In 1857 he received the French patent for his invention of the phonautograph. The phonautograph consist of a horn to capture the sound waves, that are then propagated over a membrane to a stylus. Through the movement of the stylus the sound is inscribed as fine lines onto paper covered with lamp black. The design is inspired by the anatomy of the human ear. The horn, the membrane and the stylus are designed to simulate the ear canal, the ear drum and the ossicles. However Scott did not intend to play back his recordings. Instead his phonautograph was meant as method to analyze sound visually, inspired by the then very young technology of photography [1]. It took more than 150 years until the recordings of Scott were heard for the first time. In 2008 they were reconstructed by the First Sounds collaborative using the technology developed in the IRENE project [13].

20 years after the invention of phonautograph, in 1877 Thomas Alva Edison invented the first device that was also able to play back sound, the phonograph [15]. The first phonograph used tinfoil that was wrapped around a metallic cylinder as recording medium. The sound is engraved through a stylus that performs up and down movements into the tinfoil. Hence the sound is encoded vertically as hills and dales of the groove.

The inventions of Emile Berliner initiated the transition to flat discs as medium. In 1887 he patented the gramophone. The device is in essence a variant of the phonograph that uses a zinc disc instead of a cylinder as medium. In contrast to the vertical encoding on the cylinders the sound is engraved by horizontal movements of a stylus on the disc, i.e. the audio information is encoded horizontally. The flat discs has the advantage that they can be produced in high quantities with a lower price than wax cylinders and thus paved the way for the mass adaption of sound recording.

Over time various kinds of materials have been used for flat disc recordings. The earliest discs were made from hard rubber until around 1885 shellac compounds were introduced and became standard. In 1931 the first commercially available vinyl discs were launched [17].

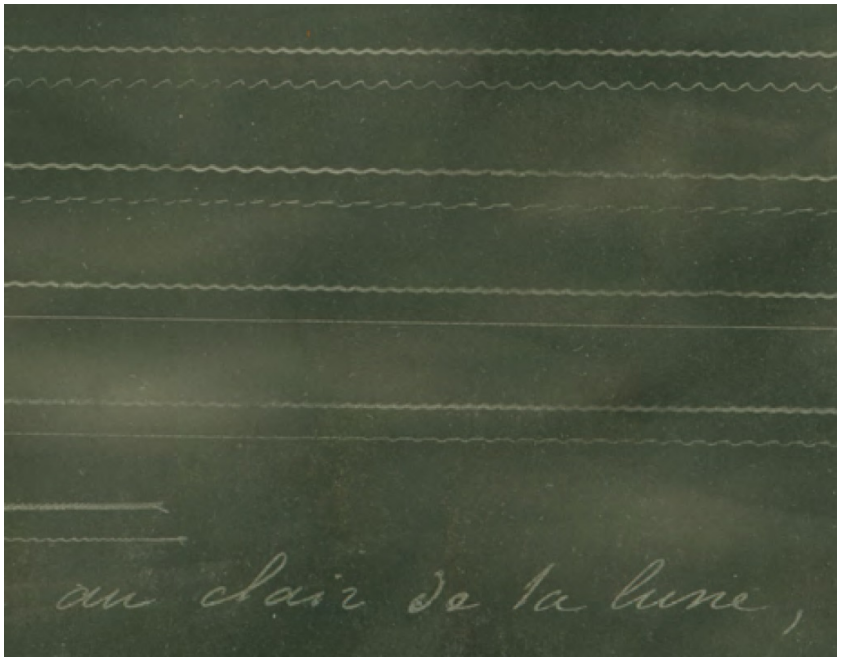


Figure 2.1.: Extract of phonautograph recording of *Au Clair de la Lune* made by Édouard-Léon Scott de Martinville in April 1860 (image source: [14])

Figure 2.2.: Thomas Edison with an early phonograph (image source: [16])

2.2. Lacquer Discs

Sound can not be recorded directly to shellac or vinyl discs as they are produced in a molding process using a stamp. Consequently there is a need for a medium on that sound can be directly recorded and from which further copies can be produced later on. Between 1930 and 1950 lacquer or acetate discs were the medium of choice for direct recordings. Before wax or aluminum discs had been used for that purpose, but they have several drawbacks. Wax discs were used for master recordings, but it was necessary to plate the disc and reproduce it in another material to get a playable copy. Aluminum discs can be played back directly after recording, but they do not offer a suitable audio quality for radio broadcasting [18]. Lacquer disc were the first medium that offered both, instantaneous playback and high sound quality. Although the lacquer is not robust enough for repeated playback, as it suffers quickly from wear through the playback stylus. Eventually the upcoming magnetic tape recordings superseded the lacquer discs as medium for direct recordings in the 1950s.

Lacquer disc consist of an aluminum plate covered with a thin layer of nitrocellulose lacquer. During the Second World War glass was used for the base plates, due to the high demand for aluminum for war efforts. The lacquer layer is so soft that sound can be recorded with a recording machine that cuts a groove with a needle, but still so solid that it can also directly be played. If additional copies were needed a metal negative, or stamper was produced in a two step process. First the lacquer disc was covered with a thin silver coating. Subsequently the stamper was created by electroplating nickel on the silver. With this stamper shellac or vinyl copies can than be pressed. The combination of instantaneous playability and reproducibility made this discs ideally suited for field or studio recordings or to archive radio broadcasts. Because of their usage many remaining lacquer disc are one-off recordings [19].

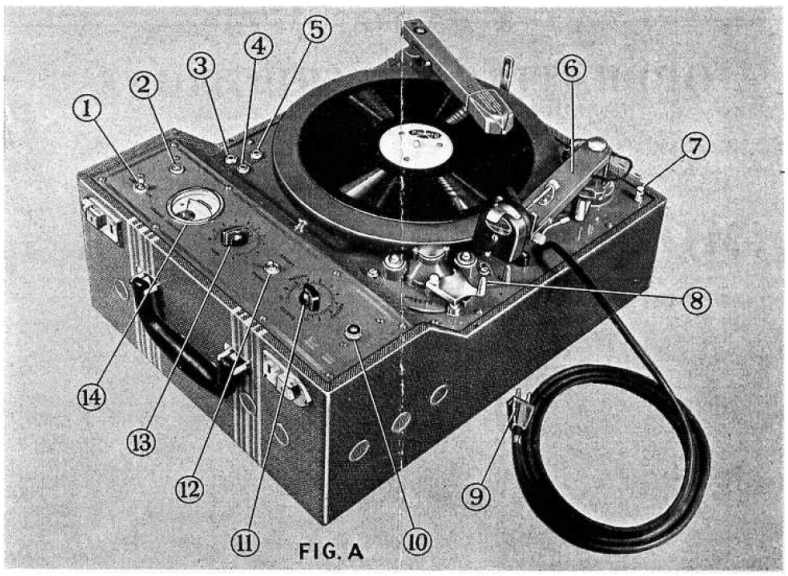


Figure 2.3.: A PRESTO Model K recording machine that was used to cut lacquer discs in the 1930s (image source: [20])

2.3. Summary

The first know sound recordings were made by French inventor Édouard-Léon Scott de Martinville in 1857, but there was no way to play them back. That changed with the invention of the

phonograph by Thomas Alva Edison twenty years later, in 1877. The first phonographs used wax cylinders as recording medium. With the advent of disc recordings around 1885 recordings became affordable for mass production.

Lacquer discs were widely used for direct recordings between 1930 and 1950. The sound can be directly recording by using a cutting machine and additional copies can be produced from the lacquer if necessary.

3. Current System

The following chapter describes the current scanning process for disc recordings. The process consists of two steps which are implemented independently. In the first step a scan of the disc surface is made with the IRENE 2D system. The system consists of the hardware setup as well as the control software and the user interface. The output of the first step are image files. For the second step, the actual audio extraction, these images are loaded into the RENE software. RENE contains several different methods for tracking of grooves which allows to adapt the audio extraction to the type and condition of the scanned disc.

3.1. IRENE 2D

3.1.1. Hardware

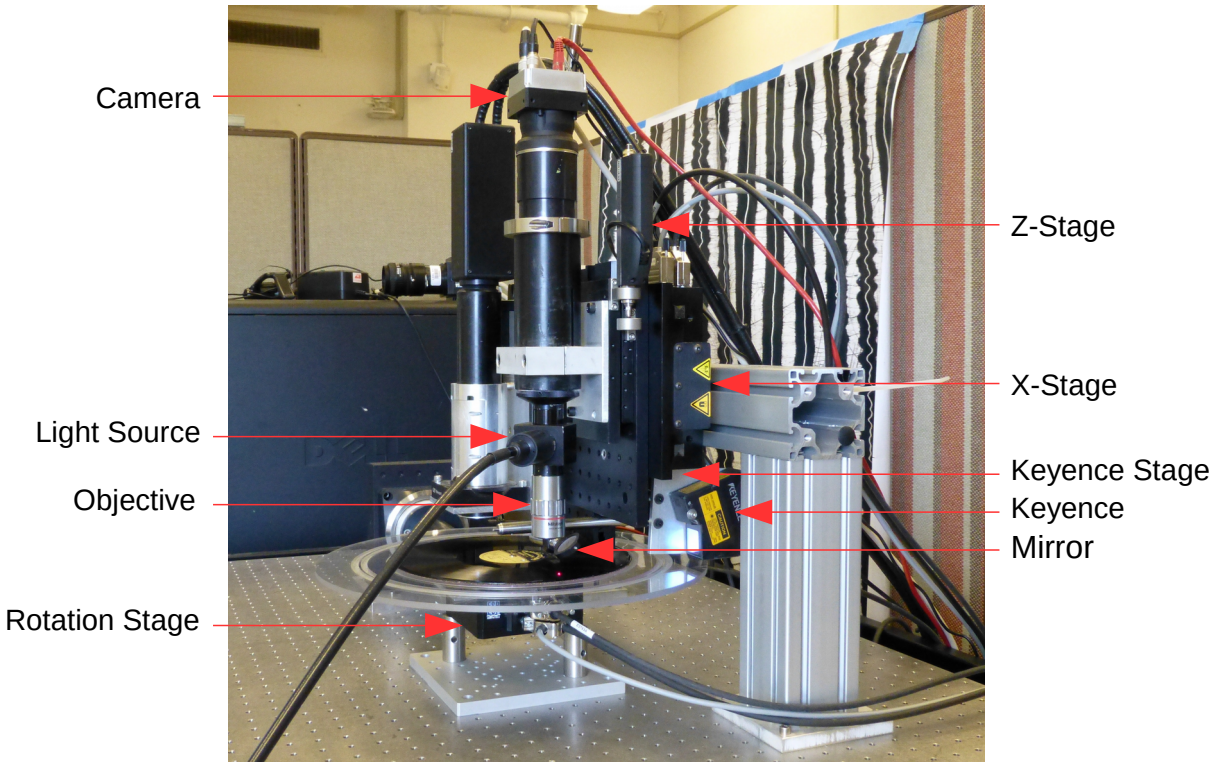


Figure 3.1.: IRENE hardware setup

The main parts of the IRENE hardware are a turntable where the disc is placed on, a line scan camera and a linear axis to move the camera over the disc during the scan. Figure 3.1 gives an overview over the current IRENE hardware setup at LBNL. The X-stage is used to move the camera over the disc during the scan. The Z-stage is necessary to adjust the distance

between the camera and the disc surface to maintain focus while scanning. The Keyence laser displacement sensor together with a mirror is used to measure the distance between camera and disc surface. The Keyence stage allows to compensate for the offset between the camera and the point where the displacement is measured (for a detailed description of the function of the Keyence stage see section 3.1.3).

All moving stages are controlled by a Newport XPS motion controller which is connected to the host computer through Ethernet. As host computer a generic PC that runs the LabVIEW control software can be used. Table 3.1 provides the most important characteristics of the used components.

TODO: (pictorial) schematic would be more useful

Controllers

Function	Device	Description
Motion controller	Newport XPS-Q8	Motion controller for 8 DC servo motors with additional 30 TTL inputs, 30 TTL outputs, 4x 14 Bit ADC and 4x 16 Bit DAC [21]

Actuators

Function	Device	Description
X-stage	Newport XML210	Linear motor stage with 210 mm travel, 1 nm resolution and 3 µm accuracy [22]
Z-stage	Newport LTA-HS	Motorized actuator with 50 mm travel, 0.1 µm resolution and 15 µm accuracy [23]
Rotation stage	Newport RV100BL	Rotation stage with 0.0001° resolution and 0.01° accuracy [24]
Displacement sensor stage	Newport LTA-HS	Motorized actuator with 50 mm travel, 0.1 µm and 15 µm accuracy [23]
Light source	Lumencor SOLA SE light engine	LED light source that produces white light within a wavelength range of 380 nm to 680 nm. Maximal intensity 3 W on a circle of 3 mm diameter [25]

Sensors

Function	Device	Description
Displacement sensor	Keyence LKG-157	CCD laser displacement sensor with a range of ± 40 mm and 0.5 µm accuracy [26]
Line scan camera	Basler Racer ral4096-24gm	Line scan camera with 4096 x 1 Pixel, Mono, 12 Bit resolution and a Pixel size of 7 µm x 7 µm [27]

Table 3.1.: IRENE hardware components for 2D scanning

3.1.2. Software

To control the whole acquisition process the LabVIEW based IRENE 2D scanning software is used. It contains all necessary software to control the moving stages, the displacement sensor and the image acquisition. IRENE 2D provides a GUI that allows the user to enter a description of the scanned disc as well as to adjust parameters like the disc size, the number of samples per revolution, the exposure time of the camera, the lighting intensity etc.

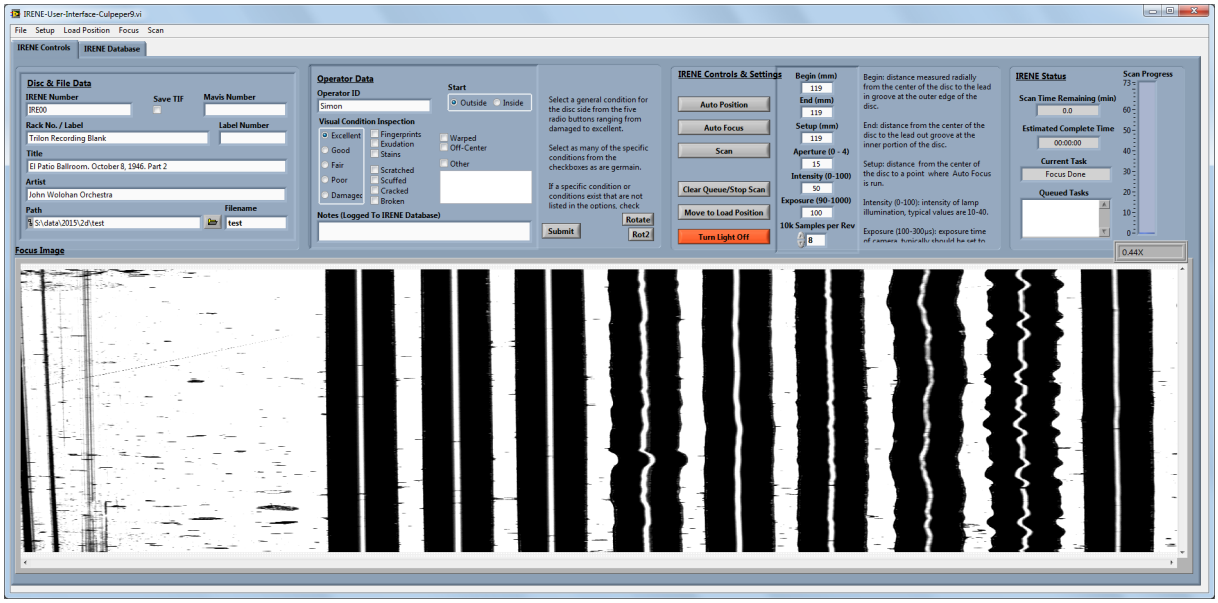


Figure 3.2.: Screenshot of the IRENE user interface

3.1.3. Autofocus

The groove bottom on a lacquer disc has a size of around 10 µm. Thus an objective with a high magnification is necessary to provide detailed images of the groove. The downside of such an objective is the limited depth of field of only 28 µm [6]. That means that the distance between the camera and the disc should not vary more than 28 µm during the scan, otherwise the image will be blurry. As the disks can not be expected to be even on that scale a method to maintain the focus during the scan is needed.

The distance between the camera and the disc surface for optimal focus is determined before the scan with an autofocusing function. This function acquires a scan of 10'000 lines over an angle of 45°. During this scan the camera is moved over a distance of 0.5 mm on the Z-axis. Afterwards the focus is estimated for every row of the image with the focus measure f_{IRENE} defined by equation 3.1. Δ_2 stands for the first order finite difference with a step size of 2 of the gray value $g(x,y)$ of a pixel. This finite difference can be seen as indicator of sudden illumination changes in horizontal direction like the ones that are created by sharp groove edges. μ stands for the average of the finite difference over the line. Hence the focus measure can be described as the standard deviation of the squared first order difference of the gray values.

$$f_{IRENE} = \sqrt{\frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N}\Delta_2^2(x,y) - \mu)^2} \tag{3.1}$$

$$\mu = \frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N}\delta_2^2(x,y) \tag{3.2}$$

$$\Delta_2 = g(x,y) - g(x,y+2) \tag{3.3}$$

After the calculation of the focus measure the line with the highest measure is determined. The optimal focus distance is the height offset corresponding to this line.

The drawback of this relative simple method is that it relies on an initial guess for the distance

around which the optimal focus is then searched. This leads to situations where the autofocusing function has to be run several times because the initial guess is to far off, e.g. after changes in the hardware setup or the disc.

To maintain the focus during the scan an active focusing mechanism with a laser distance sensor is used together with a linear stage to move the camera up and down. The Keyence laser displacement sensor continuously determines the distance between the camera and the disc and the camera position is controlled from that input.

The optimal position to measure the distance between camera and disc surface would be directly at the imaging point. This is not possible, because the laser would interfere with the image. Further it is not possible to mount the camera and the laser sensor close together, due to their size. Instead a mirror is used to deflect the laser beam, to bring the point where the distance is measured as close as possible to the imaged region. To compensate for the remaining offset in Y-direction a linear stage is used, which moves the Keyence sensor and the mirror relative to the X-stage. With every change of the scanning radius the position of the Keyence is adjusted, so that the radius of the displacement sampling point and the imaging point remain equal (see figure 3.3).

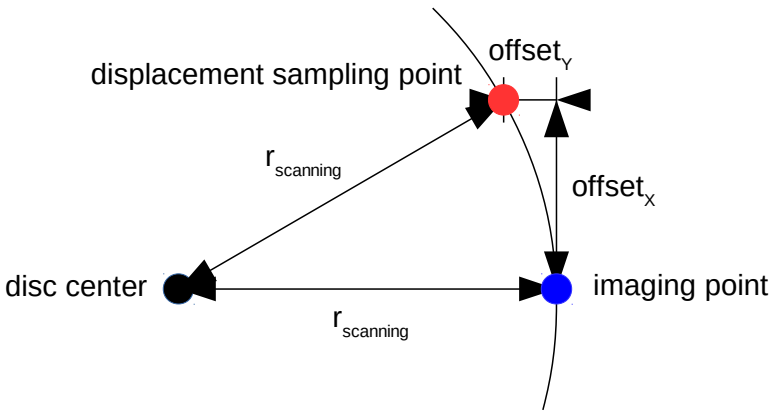


Figure 3.3.: Illustration of the adjustment of the Keyence position relative to the current scanning radius. From the current radius and the Y-offset of the displacement measurement point the X-offset is determined by $offset_x = r_{scanning} - \sqrt{r_{scanning}^2 - offset_y^2}$

3.1.4. Trigger and Resolution

The camera is triggered directly through pulses generated by the rotary encoder of the rotational stage. In this way it is possible to guarantee that the images are taken with a constant angular distance, even if the rotation speed should fluctuate.

The trigger spacing, respectively the number of triggers per revolution is set depending on the recording speed of the disc. Common recording speeds of disc are $33^1/_3$ rpm, 45 rpm and 78 rpm, but also other speeds were used. To achieve a decent audio quality and to make the scans future proof for archiving the sampling rate is usually chosen to be 96 kHz or even higher [1]. Table 3.2 contains the sampling rates that are used by the IRENE system depending on the disc recording speed. It can be seen that the sampling rate varies, because the number of triggers per revolution can not be chosen arbitrarily. The rotary encoder provides 300'000 pulses

[1] That is about twice the sampling rate of an audio CD of 44.1 kHz [28]

per revolution. The trigger spacing is then derived from these pulses by integer multiplication and division factors, which restricts the available sampling rates. Usually the closest available approximation of 96 kHz is chosen.

Recording Speed	**Number of triggers per revolution**	**Trigger spacing**	**Sampling Frequency**
$33^1/_3$ rpm	200'000	0.0018°	111.1 kHz
45 rpm	120'000	0,003°	90 kHz
78 rpm	80'000	0.0045°	104 kHz

Table 3.2.: Sampling rates of the IRENE system depending on the disc recording speed

3.1.5. Illumination

To illuminate the disc surface a LED light source is used. The light is fed into the optical system via an optical fiber and a semitransparent mirror. This results in a very concentrated lighting that is perpendicular to the disc surface as show in figure 3.4. When the light hits the even parts of the disc surface it gets reflected back to the lens. On the contrary the light that hits sloped parts of the surface gets reflected sideways and does not reach the lens. Therefore the even groove top and bottom appear bright on the image and the slopes in between are dark. The resulting hard edges in the image make it easy to track the course of the groove.

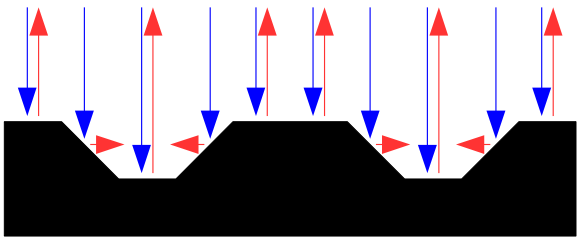


Figure 3.4.: Schematic representation of the light emitted by the source (blue arrows) and the light reflected by the disc (red arrows)

Figure 3.5.: Section of a scan of the MC101-3 laquer disc

Scanning Process and Mapping

The output of the line scan camera are images with 4096 pixel width and 1 pixel height. Once a full revolution is completed these line scans are saved in an image, where every row represents one line. Through this process a circular ring like portion of the disc surface gets mapped to a rectangular image. Hence the groove that forms a spiral on the disc appears in parallel sections on the image. If the groove was perfectly centered on the disc the groove sections on the scan would appear perfectly straight. As this is usually not the case the groove contains a slight curvature that is caused by the misalignment. This distortion does normally not affect the audio as it has a very low frequency compared to the audio information. Figure 3.6 illustrates the mapping of the groove. After a complete revolution the camera is moved in X-direction towards the center of the disc and the acquisition proceeds. These steps are repeated until the whole groove is scanned. To avoid losing parts of the groove, because they are cropped by the image border the steps in X-direction are chosen a bit smaller than the actual image size. This creates a slight overlap of the scans.

Figure 3.6.: Groove mapping

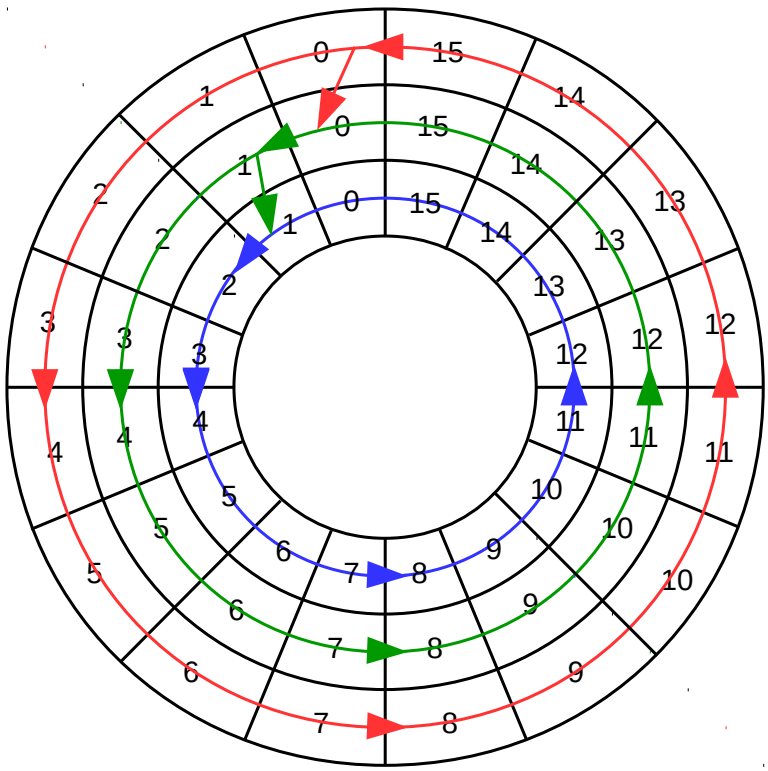


Figure 3.7.: Camera trajectory during three revolutions
(red: 1st revolution, green: 2nd revolution, blue: 3rd revolution)

The rotation of the disc is never stopped throughout the whole scanning process and the camera is continuously triggered. That also means that images are acquired during the movement of the X-Stage which are not usable for the audio extraction. The line scans are usually stored in blocks of 5000 as this is the buffer size of the camera. To compensate for the unusable lines one extra block is added during the rotations to create a time slot during which the X-stage can move. This extra block is discarded later before the image of the whole rotation is saved. Adding an extra block also means that the angle where the X-stage move changes with every rotation and therefore the starting point of the images do not match. This has to be corrected by reordering the blocks before the image of one rotation is saved. Figure 3.7 shows the camera trajectory during three revolutions.

3.2. RENE

3.2.1. Implementation

RENE is implemented in Visual C# and uses the Windows Forms library for the GUI components. Despite the object oriented nature of C# the software is written in a procedural way. Furthermore very little encapsulation is applied, most of the code is contained in a single file which has over 9'000 lines of code. For this reasons the code is rather difficult to maintain.

TODO: Add class diagram

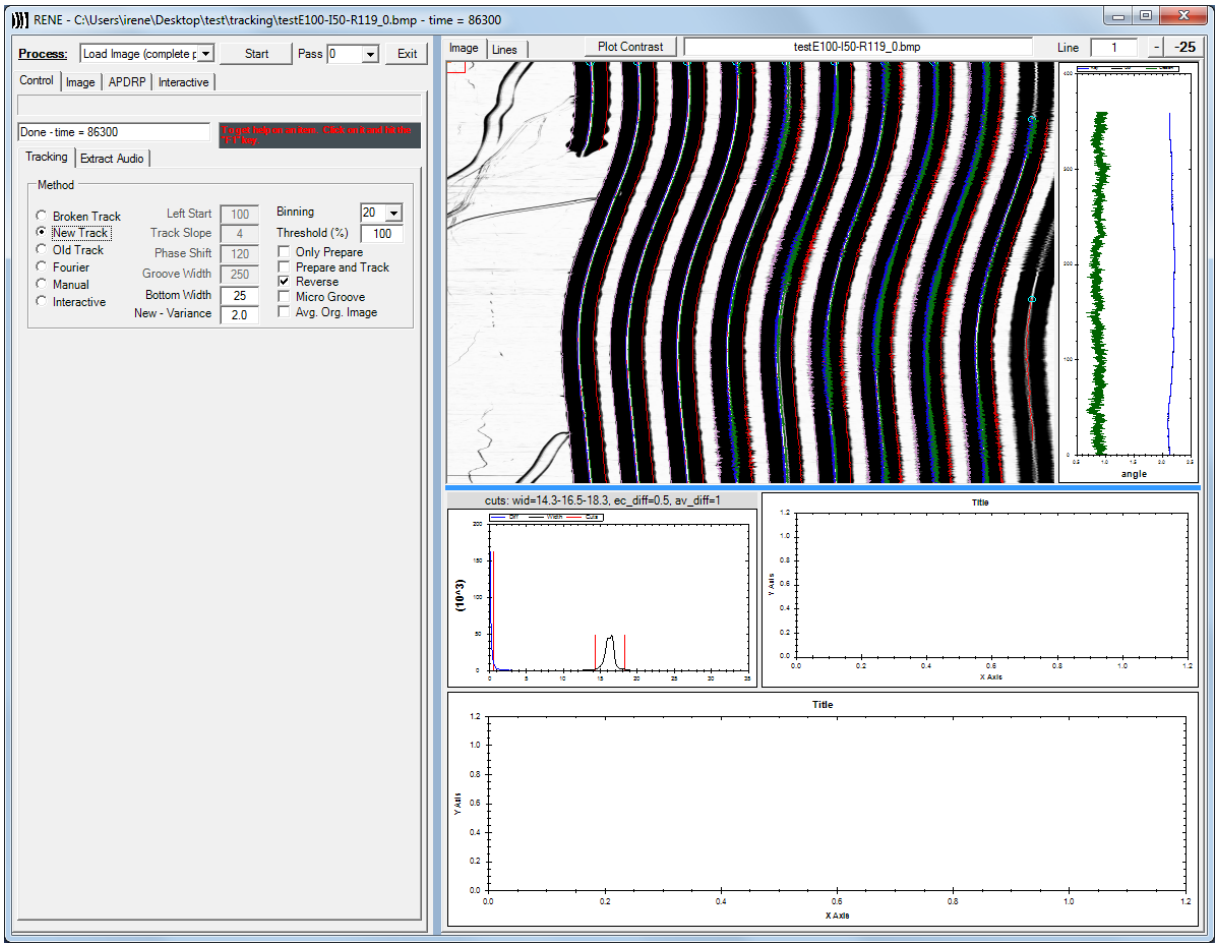


Figure 3.8.: Screenshot of the RENE user interface

3.2.2. Processing Flow

The audio extraction happens in two steps. In the first step the image is downsampled in vertical direction, usually by a factor of 20. On this binned version of the image the course of the groove is then roughly determined. This first steps is necessary to reduce the search area for the following audio extraction.

In the second step the position of the groove is determined exactly on the full resolution image. From the groove position the trajectory of a virtual stylus is then determined. The audio information can then be gained through a numerical differentiation of the trajectory. At last filters can be applied to improve the quality of the audio signal before it gets stored as WAV-File.

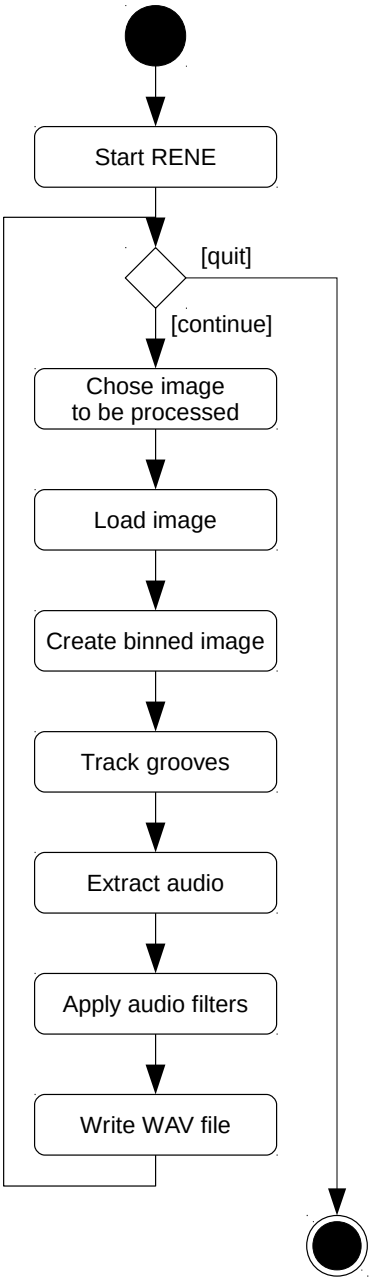


Figure 3.9.: Processing steps of the audio extraction in RENE

3.2.3. Tracking Methods

A variety of tracking methods are included in RENE. Most of them are aimed at specific types of recordings or conditions.

Old Track

The simplest tracking method named *Old Track* works by scanning the image in horizontal direction. The grooves are identified by searching for local maxima on each line which correspond to the white groove bottom. This method is based on the assumption that the grooves are continuous, hence it fails if a crack occurs.

New Track

The *New Track* improves the *Old Track* method through following the tracks from top to bottom of the image. First a start position is determined by searching for the local maxima on a line. Then the groove is tracked by searching locally for the brightest point in downward direction. If the bottom is reached the tracking starts again a the top of the image. Small breaks of the groove, e.g. by dust particles, can be linearly interpolated, but the method fails as well if a crack occurs.

Fourier

The *Fourier* method uses a Fourier transform of the image to determine the distance between the grooves on the image. The grooves on a disc normally have a constant spacing which creates a nearly periodic intensity distribution in horizontal direction. This pattern can be identified by searching for the maximum frequency in the spectrum. The spacing of the grooves is then estimated from the phase information of the maximum frequency. This method is usually very robust but fails for the start and ending parts of the disc, when parts of the image are without a groove. The method fails as well if a crack causes a horizontal shift in the groove.

Manual

The *Manual* tracking method relies on the user to track the groove and enter points on the groove which are then connected by straight lines. It serves as backup method for especially difficult cases when all other methods fail.

Interactive

The *Interactive* method was developed by J. Singy. The basic idea is to increase the efficiency of the manual tracking while keeping the flexibility. For this purpose a magnified section of the image is moved vertically over the image, similar to scrolling over an zoomed image. During the process the user enters the horizontal position of the groove with the mouse. From this point the exact position of the groove center is then determined by the software [9].

Broken Track

This method was developed by S. Fischer and is especially targeted to relink tracks across breaks. The method is based on a Canny edge detection to identify the edges at the top and bottom of the groove. The edges are then tracked from top to bottom of the image. In the next step grooves are the relinked automatically. For that purpose a Fourier transform of the groove is made and the spectra of both ends are compared. The assumption herein is that that the audio signal has a nearly constant spectrum for short periods of time. Therefore the ends with the highest similarity in the spectrum are also the most likely to have been connected. In cases where the method fails, e.g. because a high number of cracks are present, the cracks can also be manually relinked [7].

3.3. Summary

The workflow for scanning disc recordings is currently implemented in two distinct steps. The scanning of the disc is done with the IRENE 2D setup, which includes the scanning machine as well as the control software. The output of the first steps are images of the disc surface which are the feed into the RENE software for audio extraction.

Part II.

Image Acquisition on Cracked and Delaminated Discs

4. IRENE 2D Upgrade

Silvan Fischer developed a acquisition method that allows to get scans with good focus even when cracks in the disc surface occur [7]. The acquisition method named *Crack Handler* detects the cracks by analyzing the output of the laser displacement sensor. If a crack is detected an additional scanning round at the same radius is made. For this round the camera trajectory is adapted to get the blurred portions of the previous scan in focus.

Unfortunately was the existing IRENE 2D software upgraded while Fischer worked on his thesis. The previously used Dalsa Piranha P2-42-04K40 line scanning camera [29] was replaced by a Basler racer ral4096-24gm which provides superior image quality [27] [1]. This leads to the situation that the *Crack Handler* method can not be used with up to date hardware and thus was not in productive use since it's implementation. Further is a fair comparison of the *Crack Handler* with any new methods impossible at the current state of the acquisition software. Therefore an upgrade of the *Crack Handler* method is necessary.

4.1. Feature Assessment

To perform the update in an efficient manner the differences between the two versions of the IRENE 2D control software need to be identified. The differences in functionality will be grouped in features which then can be divided into necessary and compulsory features. The focus of the upgrade will be on the features which are necessary for a comparison with the to be developed methods.

All differences can be found in the IRENE-Control-PX.vi file which holds the main control sequence for the acquisition. The file contains two version of the acquisition control, the one implemented by Fischer which detects cracks on the fly and the standard method. The following sections describe the main differences between the two versions.

4.1.1. Camera API

The previously used Dalsa camera uses a camera link interface for the data transfer which requires a dedicated interface card on the host computer [29]. The new Basler camera uses a GigE vision interface that can be connected to a standard Gigabit Ethernet port [27]. The drivers and interfacing VIs for both cameras are both part of the NI Vision Acquisition Software, but the cameras need different drivers. Camera link can only be interfaced from the IMAQ driver, while the GigE interface needs the IMAQ-dx driver [30]. Therefore the interfacing VIs need to

[1] The problem with the image quality of the Dalsa camera is caused by the use of two ADC taps to read out the CMOS chip [29]. That means that every other pixel is connected two one of two separate ADC channels. This configuration is chosen to speed up the readout process. Unfortunately the two channels never have the exact same gain. That leads to a slight difference in brightness between two adjacent pixels, which is visible as a stripe pattern in the image. The difference in brightness can be compensated by calibrating the gains of both stages in the camera software. However the calibration could never be made exact enough to cancel out the stripe pattern. Therefore it was decided to replace the camera.

be exchanged. The drivers differ slightly in their functionality, so the exchange is not just a drop-in replacement and a few adaptions have to be made.

4.1.2. One Image per Revolution

Previous version of IRENE 2D stored the scans in images of 10'000 lines, due to memory allocation limitations in older Windows versions. The exact number of 10'000 was chosen because it is the maximal buffer size of the Dalsa camera. One scanning revolution can have between 80'000 and 200'000 lines, i.e. the scan for one revolution was divided into between 8 and 20 sectors.

Figure 4.1.: Division of the scan into sectors (not to scale)

The number of images per revolution needed to be changed with the upgrade to the new camera. The Basler camera has a smaller maximum buffer size of 7147 lines. As storing even more images per revolution is impractical and the memory allocation issue has been overcome with new Windows versions the image buffering was redesigned. The lines are now buffered in chunks of 5'000 and then stored as a single image per revolution. 5000 lines was chosen because it equals half the previous buffer size and made the upgrade of the software easier. The new way to handle the images also made the reordering of the buffers necessary, which is explained in section 3.1.5.

4.1.3. Motorized Keyence

The linear stage that is used to adapt the offset between the camera and the laser displacement sensor was added to the IRENE setup after Fischer completed his thesis. The stage is necessary to keep the point where the displacement between disc surface and camera is sampled and the position of the camera on the same radius relative to the disc center (for a detailed description of the working principle see section 3.1.3).

In the previous version of IRENE the displacement sensor had a fixed offset relative to the camera. Consequently the displacement is not sampled at exactly the same radius as the image

is taken. Instead the sensor is placed, so that it is one or several steps in X-direction inward of the camera position (see figure 4.2). This way the displacement of the disc is sampled at least one revolution ahead and is then stored in a buffer until it is needed for the adjustment of the camera height.

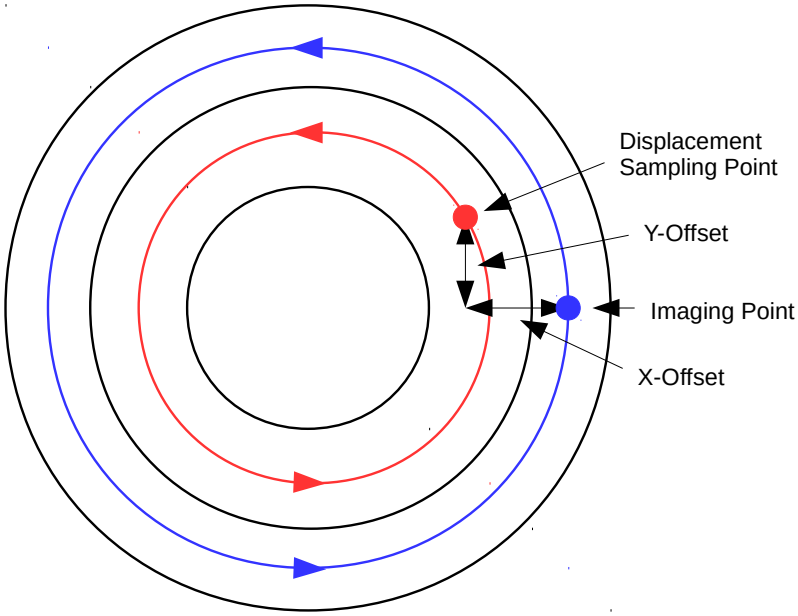


Figure 4.2.: Displacement sensor with fixed X-Offset

4.1.4. TIFF Format

The scans from IRENE where usually stored as bitmaps in the BMP format. Recently the option to store TIFF images was added. TIFF was chosen because it allows to add metadata to the image itself. This way information about the scan or the disc itself, e.g. illumination intensity, the exposure time or the general condition of the disc, can be stored together with the image data.

4.1.5. Categorization of Features

The exchange of the camera API is obviously necessary to run the new camera, so it has to be considered a necessary feature.

As the buffer size of the new camera is lower than the previous one, the way the images are saved has to be adapted in some way. Going with 5000 lines images does not make sense, because it would double the amount of pictures per revolution and make handling the scans more cumbersome and error prone. Therefore an update to one single image per revolution is the way to go.

Although a new linear stage was added, the setup can still be run the old way with a constant offset between Keyence and camera. The motorized Keyence could increase the accuracy of the crack detection slightly, because it eliminates the error between the displacement sampling and the imaging position. For the comparison with new methods this upgrade is not absolutely necessary, as the improvement is probably low. Hence the motorization of the Keyence is

considered a complementary feature.

The switch to TIFF files does not change the scanning process in any way. Consequently it is not necessary for comparison of the methods and is considered a complementary feature.

4.2. Implementation

Because the Basler camera is already used in the standard scanning process, no new VI to interact with hardware have to be designed. This makes the upgrade a straightforward process. Both necessary upgrades, the new camera API and using only one image per revolution are implemented.

4.3. Test

4.3.1. Test Setup

To make sure that the upgraded system works as intended, a broken disc is scanned. Because Fischer broke the discs gradually during the tests a direct comparison is impossible. Therefore another disc has to be used. Nevertheless, if a cracked disc can be scanned and the audio extracted it can be assumed that the system works. For the test a shellac disc in good condition with a single lateral crack is used (see figure **??**). Three sections of the disc are scanned, the outermost, the middle and the innermost part.

TODO: Add picture of the disc

4.3.2. Results

The disc could successfully be scanned. The lateral crack gets detected and the camera trajectory is adapted for the second revolution in all sections.

The new scans with only a single image per revolution can be loaded into RENE without any problems. The automatic re-linking of the grooves with the *BrokenTrack* method fails for this disc. *BrokenTrack* has a range of parameter that can be adjusted to improve the results. However, no working setting could be found for any section of the disc. Thus the audio extraction can only be done through manual tracking of the grooves. Although a audio extraction with *BrokenTrack* was not possible, the test shows that the upgraded acquisition system works as intended.

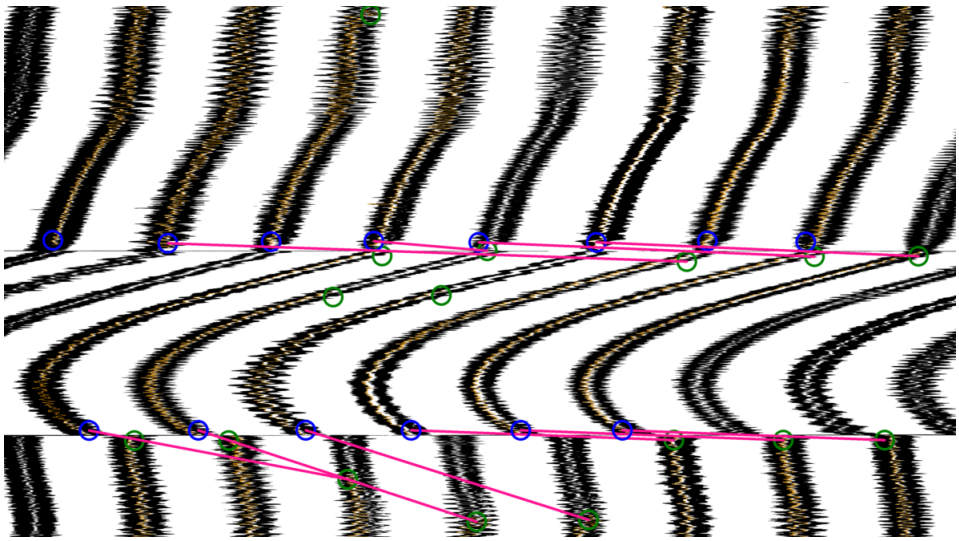


Figure 4.3.: Output of the automatic re-linking of broken tracks with the *BrokenTrack* method. One can see that the automatic re-linking of the groove ends (green) and beginnings (blue) produces inconsistent results.

4.4. Summary

In the first step of the upgrade process the differences between the current acquisition software and the one implemented by Fischer are identified. In the following the features necessary to use Fischer's acquisition method with the new Basler line scanning camera are implemented. Finally the upgraded acquisition method was tested successfully through scanning of a broken shellac disc.

5. Focus and Depth of Field

To determine which spacing should be used for the acquisition of focus stacked images the depth of field of the image acquisition system has to be determined. The first section of this chapter presents the theoretical relationship between the parameters of the optical system and the depth of field and contains a theoretical estimation. The second part presents a serious of measurements to determine the depth of field of the IRENE system.

5.1. Theoretical Estimation of the Depth of Field

The spatial resolution as well a the depth of field of a microscope can be estimated by calculating the dimensions of the airy disc. The airy disc is the disc on which a point in the image is projected on, caused by the diffraction of light. The height of the disc Z can be calculated as in equation 5.3. n is the refraction index of the medium, in the case of air n equals 1. λ is the wave length of the light used and NA is the numerical aperture. The numerical aperture depends on the half angle of the cone from which the objective collects light θ and the refraction index n (see equation 5.2) [31]. The currently used Mitutoyo objective has a numerical aperture NA of 0.14 [6]. The Lumencor light engine produces white light that covers the range from 420 nm to 680 nm [25]. Thus the wavelength λ is assumed in the middle of this range as 550 nm. Using this values Z equals 28 µm.

$$Z = \frac{n\lambda}{NA^2} = \frac{1 \cdot 550\ nm}{0.14^2} = 28\ \mu m \tag{5.1}$$

$$NA = n\ sin\theta \tag{5.2}$$

The estimation based on the airy disc provides a lower boundary of the depth of field of the system, as it does not consider the resolution of the camera. The spatial resolution of the camera is usually much bigger than the diameter of the airy disc, therefore some blur of the image acceptable. In this case Berek's formula as given by equation 8.4 can be used to estimate the depth of field [32]. M is the magnification of the system, in this case 5. e is the smallest distance that can be resolved by the detector, which equals to the pixel size of 7 µm in this case [27]. Using the Berek's formula the depth of field of the system is estimated as 38 µm.

$$Z = \frac{n\lambda}{NA^2} + \frac{n}{MNA}e = \frac{1 \cdot 550\ nm}{0.14^2} + \frac{1}{5 \cdot 0.14}7\ \mu m = 38\ \mu m \tag{5.3}$$

5.2. Focus versus Added Noise

The purely optical estimation of the depth of field of the system does not provide any information about how the audio extraction is influenced by the focus of the images. To get an idea of how an offset from the best focus plane affects the audio an experiment is carried out.

The lead in groove of a lacquer disc is scanned with the current IRENE 2D system and then

the sound is extracted from the scan. At first a scan at the best focus distance from the disk is taken. Afterwards scans with both positive and negative height offsets from the best focus are made. The lead in groove does not contain any sound information, therefore the extracted sound can be seen as indication of the inherent background noise level introduced through the imperfections of the disc and the audio extraction process. As the images of the grooves get more and more blurry the probability of errors in the tracking process increases. Consequently an increase of the overall noise level is expected. Besides that, an attenuation of the higher frequency parts is expected, because the blurring of the image eliminates small features of the groove.

A lacquer disc named MC-101-3 is used as test object. It has been used for similar investigations before, because of its very good condition. Consequently the noise level from imperfections of the disc, e.g. dust or scratches, should be minimal. The best focus distance from the disc is determined with the fast focus function currently used in the IRENE 2D software (for a detailed description of the focus measure see section 3.1.3). Figure 5.2 shows a section of a scan with the start of the lead in groove on the left. The lead in section on the left appears completely straight on the scan in contrast to the sections on the right containing audio information.

Figure 5.1.: MC-101-3 lacquer disc

To quantify the noise the signal RMS is taken over the whole frequency range of 0 Hz to 52 kHz. Only the silent part of the extracted sound, corresponding to the lead in groove, is used for comparison. This way the spectrum is not biased through audio information. The audio analysis is done in a Python script using the SciPy library for scientific computing [33] including the modules NumPy for numerical calculations [34] and Matplotlib for graphical presentation [35].

Figure 5.3 shows the plot of the signal RMS in dB versus the offset in Z-direction. The RMS of the signal is calculated for the first 2.4 s of the extracted audio files, which contain no audio information. Between the best focus position and an offset of 20 µm the RMS stays roughly the same. In the range between −10 µm and 40 µm the RMS increases less than 1 dB.

Figure 5.2.: Lead in groove on the MC-101-3 lacquer disc

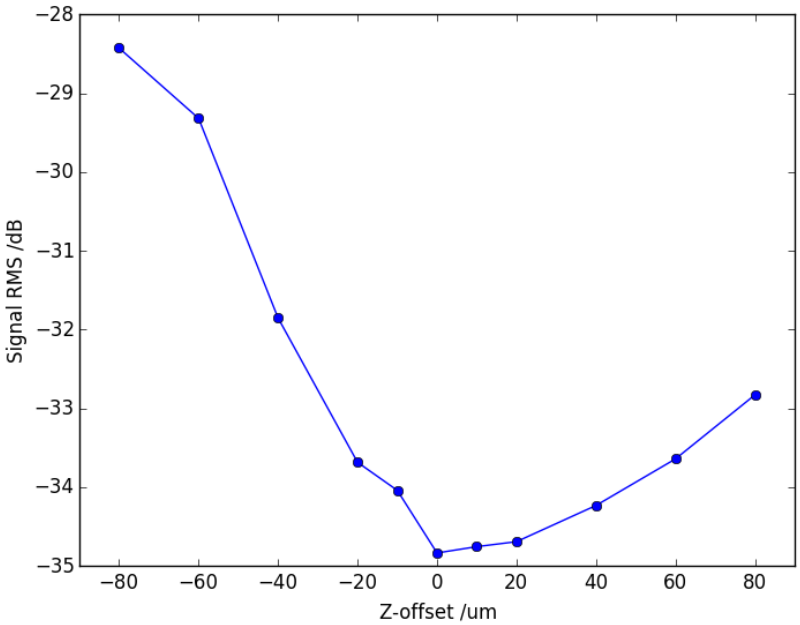


Figure 5.3.: Signal RMS of the lead in groove versus Z-offset from the best focus plane

Figures 5.4 and 5.5 show the waveform and spectrum for the Z-offsets of 0 µm, 80 µm and −80 µm. The waveforms are normalized, so that they fit in the range of 0.0 to 1.0. To calculate

the spectrum the audio files is splitted in chunks of 8192 samples over which a fast Fourier Transform (FFT) is calculated. The spectrum for the whole 2.4 s is then calculated by averaging the spectra of the chunks. The frequency resolution of the FFT is $f_r = \frac{f_s}{n} = \frac{104kHz}{8192} = 12.7Hz$, with f_s representing the sampling frequency. The spectrum shows that most of the added noise occurs in the range 1 kHz. Around 20 kHz a dampening effect seems to occur that could be caused through the blurring of the image, but this is speculative.

Figure 5.6 shows a comparison of the same section of the scans over different Z-offset. The first two images with a negative offset appear blurred overall, where as the image with best focus in the middle shows sharp edged on the groove bottom. The last two images with a positive offset show less detail at the groove bottom but the groove top seems to be in better focus.

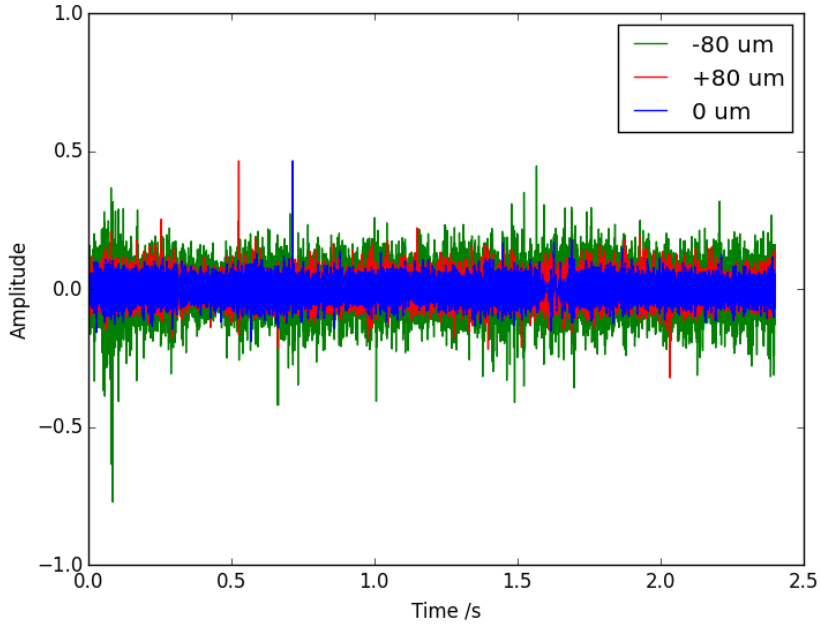


Figure 5.4.: Signal waveform of the lead in groove for 0 µm, 80 µm, −80 µm offset from the best focus plane

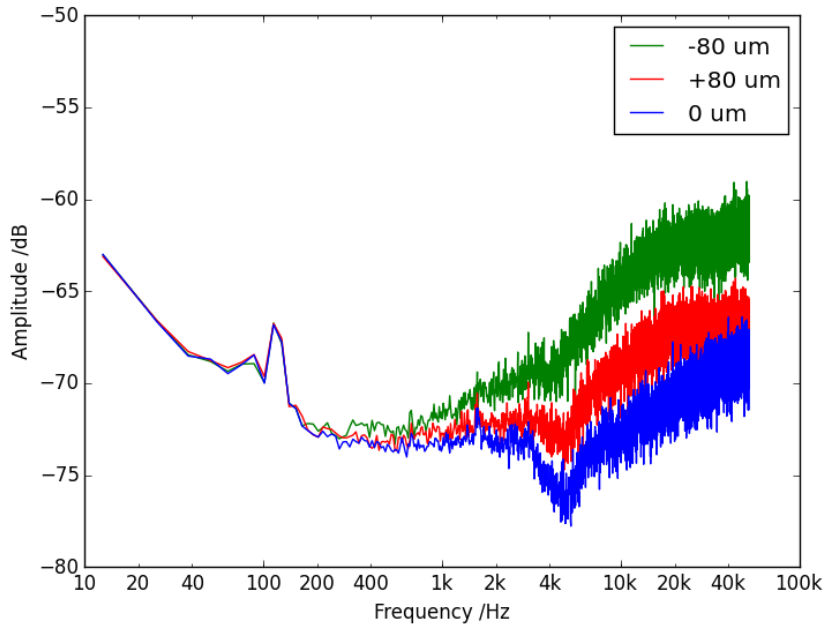


Figure 5.5.: Signal spectrum of the lead in groove for 0 µm, 80 µm, −80 µm offset from the best focus plane

5.2.1. Aperture

As it can be seen from equation 5.3 the depth of field is inversely proportional to the squared numerical aperture. Consequently it should be possible to increase the depth of field the system by decreasing the size of the aperture. This relation becomes obvious on a pinhole camera which has theoretically an infinite depth of field. To investigate the possibility of increasing the depth of field with a smaller aperture an aperture with variable size is added in the optical path after the objective. With this setup the depth of field experiment is repeated with different aperture settings.

A smaller aperture comes with the drawback that the amount of light that reaches the camera is reduced. This can be compensated by increasing the illumination intensity or the exposure time. Increasing the the illumination intensity can be done easily, but the approach is limited by the power of the light source. When the illumination is set to the maximum value the only option is to increase the exposure time. Currently the scanning speed of the system is mainly limited by the necessary exposure time, therefore any increase in exposure time would increase the scanning time.

Theoretically the necessary increase of the illumination intensity or the exposure time should be proportional to the decrease in aperture area. If the aperture diameter is decreased from 12 mm to 10 mm the area decreases by 40%, i.e. the necessary increase in illumination intensity should be 40% as well. In practice that does not hold true, to get images with comparable brightness the the illumination has to be increased more than expected. Therefore the illumination intensity and exposure time settings are chosen empirically through comparison of the output images. Table 5.1 shows the different settings used for every aperture diameter.

Aperture Diameter	**Illumination Intensity**	**Exposure Time**
12 mm	50%	100 µs
10 mm	100%	100 µs
8 mm	100%	200 µs

Table 5.1.: Illumination time and exposure time settings used with different aperture diameter

Figure 5.7 shows the signal RMS vs. the offset from the best focal plane for all aperture values. Reducing the aperture from 12 mm to 10 mm reduces the noise slightly but the illumination intensity has to be doubled. A further decrease of the aperture diameter does not further decrease the noise.

5.3. Conclusion

The smallest increase in amplitude of a sound that can be perceived by humans, called the just-noticeable difference (JND) is around 1 dB [36]. In the range from −10 µm to 20 µm the Signal RMS changes less than 1 dB. This would allow for a distance of 50 µm, although not symmetrical. As the value at −20 µm with 1.15 dB difference is still close to the 1 dB limit, the spacing for the scans is set to 40 µm, which seems to represent a good middle ground between signal quality and the size of the image stack. The depth of field of 40 µm is also backed by the theoretical estimation in section 5.1 which lead to 38 µm.

The high increase of the illumination intensity necessary to maintain the brightness of the image makes the additional aperture impractical. The MC-101-3 disc used for the test is in very good condition and therefore it can be scanned with a relatively low lighting intensity. However most

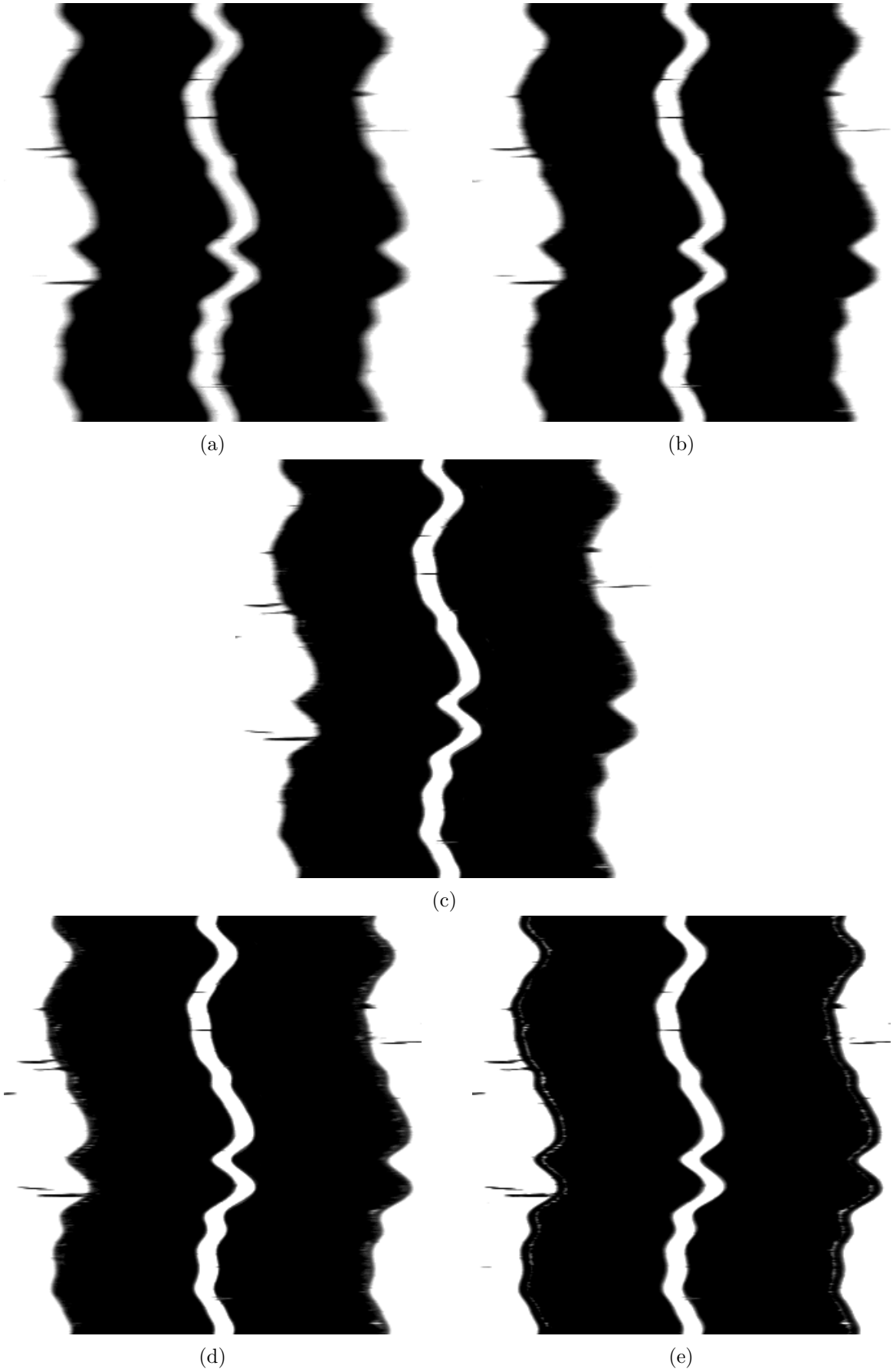


Figure 5.6.: Comparison of scans with different Z-offsets ((a) −40 µm, (b) −20 µm, (c) 0 µm, (d) 20 µm, (e) 40 µm)

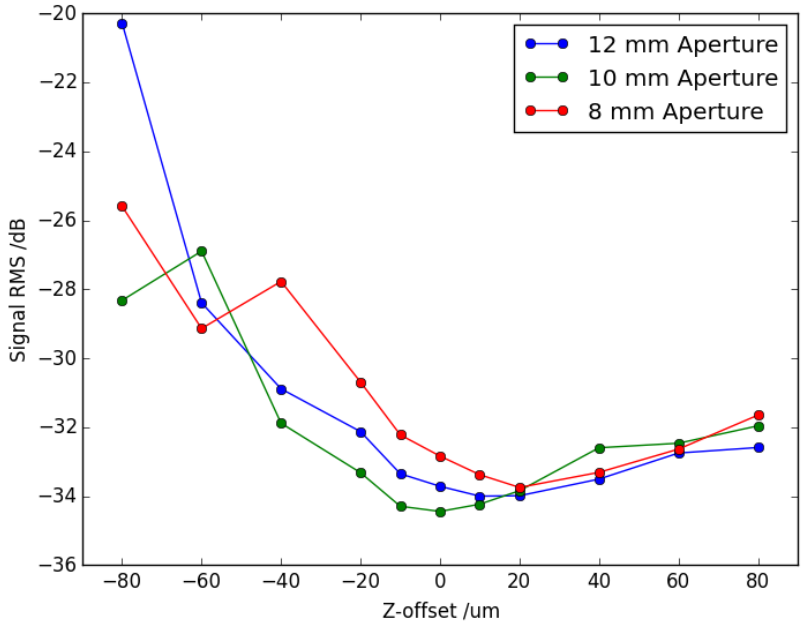


Figure 5.7.: Signal RMS of the lead in groove versus Z-offset from the best focus plane

of discs require higher settings to get good results. With a smaller aperture there would be no room to increase the intensity if the disc is in bad condition. Increasing the exposure time and hence increasing the scanning time would be the only option, which is undesirable.

5.4. Evaluation of Focus Measures

A focus measure is basically a method to caluclate the likelyhood for a pixel beeing in good focus. The experiment described in this section is carried out to get an idea how some widely used focus measures behave relative to the offset from the best focal plane. In hindsight of the image fusion the accuarcy and the selectivity of the focus measures are especially important. Accuracy means how precise the focus measure indicate the best focal plane. The selectivity is how well the measure differentiates between in focus and out of focus images.

5.4.1. Test Target

As mentioned previously lacquer disc usually have a warpage. Furthermore the distance between groove top and bottom is usually around 70 µm. Overall lacquer disc can not be considered flat and are therefore not well suited as test objects. Instead an optical test target made from glass with a test pattern consisting of fine black lines is used as test object [37]. Figure 5.8 shows the used test object.

5.4.2. Acquisition

To get a stack of scans from the test object the existing autofocus VI is modified. The stack consists of 250 images which are acquired over a scanning range of 500 µm with 2 µm spacing. Each image has a width of 4096 and a height of 1000 pixels. Figure 5.9 shows an example from

Figure 5.8.: R1L3S10P test target from Thor Labs (image source: [37])

the stack of images. The 1000 lines are acquired at a radius of about 180 mm with an angular step of 0.0045°. That leads to a pixel size of about 14 µm in vertical direction compared to about 0.5 µm in horizontal direction. Therefore the test pattern appears squeezed in vertical direction compared to image 5.8. The slight bend of the lines in the pattern is caused by the circular movement of the camera.

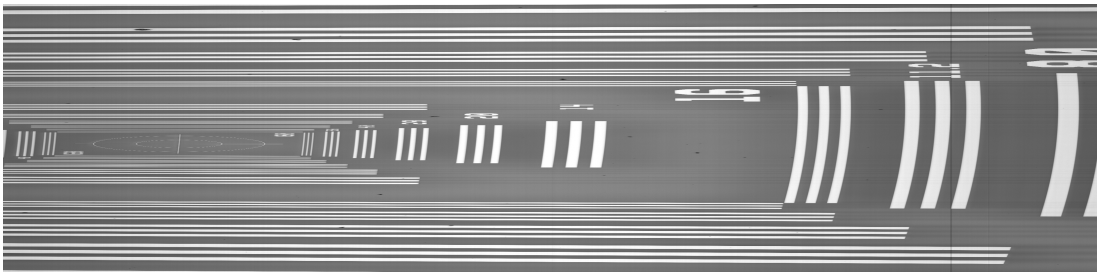


Figure 5.9.: Image taken from the R1L3S10P test target

5.4.3. Focus Measures

Based on the paper from H. Mir, P. Xu, and P. van Beek [38] several focus measures are chosen for evaluation. The focus measures can be categorized into families according to their basic working principle. The families are first order derivative, second order derivative, histogram, image statistics and correlation. The best performing measure from each family is chosen, except Vollath's F4 which is chosen on the high number of recommendations in other papers [39, 40]. The following sections describe the used focus measures.

Brenner

The Brenner measure belongs to the category of the fist order derivative based focus measures. The idea behind these measures is that a in focus image contains sharp edges. These sharp edges lead to to high derivative values which are taken as indicator of focus. The Brenner measure is based on the first order finite difference of the gray levels in horizontal direction [38]. To get the focus measure for a given image, the sum of the squared difference is calculated as in equation 5.4. $g(x,y)$ denotes the gray level of the pixel at position (x,y).

$$f_{Brenner} = \sum_{y=1}^{M} \sum_{x=1}^{N} (g(x,y) - g(x-2,y))^2 \tag{5.4}$$

Sobel

The Sobel focus measure belongs as well to the category of first order derivative based measures. It uses the well know Sobel gradient operator to identify sharp edges in the image [38]. The Sobel operator combines the differentiation with smoothing through the emphasis of the center point [41]. The operator computes first the horizontal and vertical gradients G_x and G_y through convolution of the image I with a 3 by 3 kernel (see equation 5.5, 5.6) [42].

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \tag{5.5}$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I \tag{5.6}$$

The gradient image is then obtained by calculating the magnitude as in equation 5.7.

$$G = \sqrt{G_x^2 + G_y^2} \tag{5.7}$$

After the edge detection the squared sum over all pixels is taken to calculate the energy of the gradient (see equation 5.8). In some applications a threshold is applied to the gradient image, to suppress low gradient values. These variant of the Sobel focus measure is usually referred as Tenegrad after J. M. Tenenbaum [40].

$$f_{Sobel} = \sum_{y=1}^{M} \sum_{x=1}^{N} G(x,y)^2 \tag{5.8}$$

Laplacian

The Laplacian focus measure is based on the second order derivative of an image. It can also be used as indicator of focus analog to the first order derivatives [38]. The operator can be described as the convolution of the Image with a 3 by 3 kernel as in equation 5.9 [41].

$$L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & +8 & -1 \\ -1 & -1 & -1 \end{bmatrix} * I \tag{5.9}$$

Again the focus measure is calculated as sum of the square of every pixel (see equation 5.10).

$$f_{Laplacian} = \sum_{y=1}^{M} \sum_{x=1}^{N} L(x,y)^2 \tag{5.10}$$

Histogram Range

This measure is base on the assumption that a image with good focus shows a wider spread in the histogram [38]. First the histogram $h(k)$ of the image is calculated. The focus measure is then simply the difference between the maximum and the minimum of the histogram as in

equation 5.11.

$$f_{Histogram} = max\{h(k)\} - min\{h(k)\} \tag{5.11}$$

Normalized Variance

The normalized variance belongs to the focus measures based on image statistics. The basic assumption is similar to the range of the histogram, that a sharp images shows a wider distribution of the intensity values [38]. As the name implies the measure is based on the calculation of the variance of the pixel intensities as in equation 5.12.

$$f_{Variance} = \frac{1}{MN\mu} \sum_{y=1}^{M} \sum_{x=1}^{N} (f(x,y) - \mu)^2 \tag{5.12}$$

Vollath's F4

Vollath's F4 is a focus measure based on autocorrelation. The family of correlation based focus measures is, like the derivative based ones, built on the assumption that a in focus image contains sharp edges. The sharper the edges in the image are, the narrower will be the peak of the autocorrelation. Hence the output value of the autocorrelation can be used as a focus measure. Equation 5.13 shows the calculation of Vollath's F4 measure. It uses the difference between two autocorrelations, one with a shift of 1 and one with a shift of 2. The focus measure is then the sum over these differences [38].

$$f_{Vollath\ F4} = \sum_{y=1}^{M} \sum_{x=1}^{N-1} f(x,y) f(x+1,y) - \sum_{y=1}^{M} \sum_{x=1}^{N-2} f(x,y) f(x+2,y) \tag{5.13}$$

IRENE measure

For comparison the focus measure that is currently used for the autofocusing mechanism in IRENE 2D is implemented as well. It can be seen as a combination of a first order derivative with statistical methods, as it uses the variance of the first order derivative. It is described in detail in section 3.1.3.

5.4.4. Implementaion

The focus measures are implemented in Python. To speed up the implementation the OpenCV Library [43, 44] is used for the image processing and the NumPy package for the numerical calculations [34].

5.4.5. Conclusion

Figure 5.10 shows the output values of the focus measures over the whole stack of images. With the exception of the histogram range all measures show a easily identifiable maximum and monotonic rise, respectively fall of both sides of the maximum. The histogram range does not show a clear candidate for the best focus plane at all. The Sobel measure and the variance do

not point to the same position as the other measures. The best focus position for the Sobel and Variance measure is at 4.624 mm, where as Brenner and Vollath's F4 point to 4.638 mm. That is a difference of 14 µm. The IRENE measure points to 4.640 mm, which is only a 2 µm difference to Brenner and Vollath's F4. Brenner, Vollath's F4 and the IRENE measure all have in common that they only work in horizontal direction which seems to lead to better results.

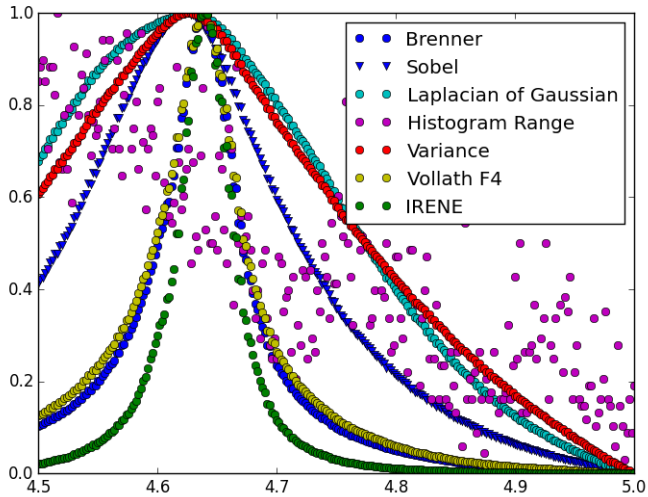


Figure 5.10.: Focus measures calculated over a stack of images taken from the test target

5.5. Summary

The depth of field of the current IRENE system can be theoretically estimated to be about 28 µm. In practical experiments the noise in the extracted audio increases significantly when the offset from the best focus distance exceeds 40 µm, respectively −10 µm. According to the measurement results the spacing of the focus stacked scans will be set to 40 µm.

The evaluation of focus measures shows that measures that work only in the horizontal direction lead to better results.

6. Acquisition of Focus Stacked Images

In the current IRENE 2D acquisition system the Z-position of the camera is controlled from the output of the displacement sensor, so that the camera stays in optimal focus distance to the disc (see figure 6.1). The optimal focus distance is determined in advance with an autofocusing function (for a detailed description of the current system see section 3.1.3).

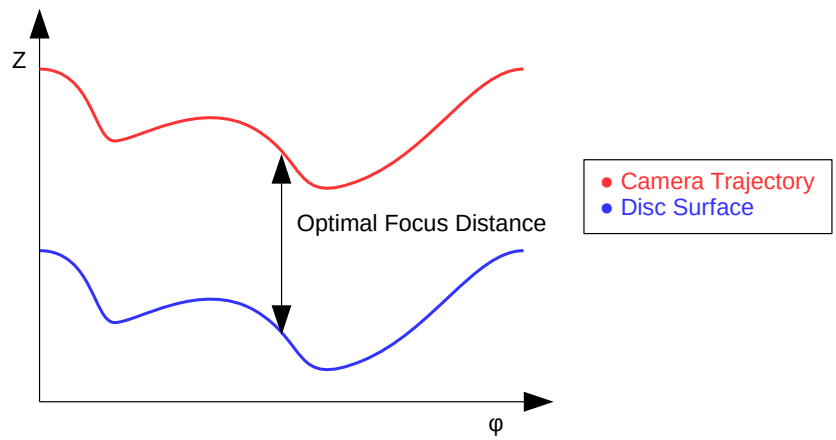


Figure 6.1.: Current acquisition method with tracking of the camera

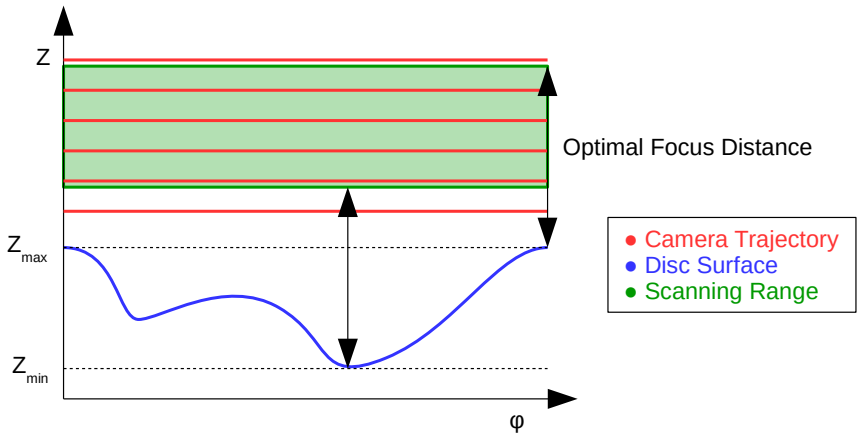


Figure 6.2.: New focus stacking acquisition method

The system performs well if the disc surface is smooth. However if a break occurs the camera position can not be adjusted fast enough, due to mechanical limitations of the linear stage and the camera weight. This results in a blurred area after the crack (see figure 6.3).

The new acquisition method aims to avoid this problem by taking a stack of images with different focal planes. The disc gets scanned over the range from minimum to maximum height with constant spacing of the scans (see figure 6.2). The idea is, that in the end every scans contains

parts of the disc surface in good focus (see figure 6.4). This stack of scans is then merged into one image with good focus everywhere in the next processing step.

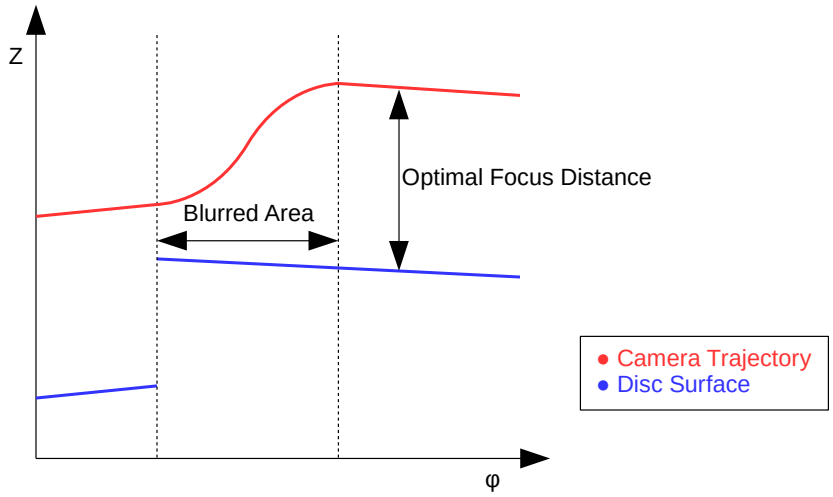


Figure 6.3.: Behaviour of the tracking method when a crack occurs

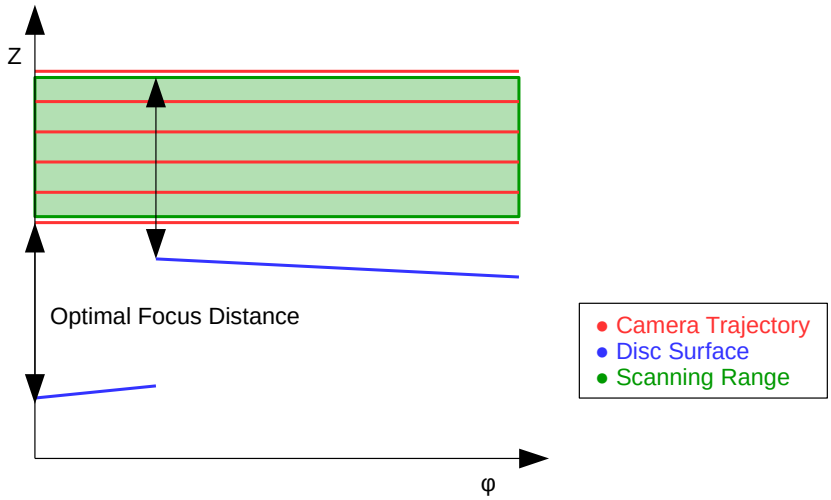


Figure 6.4.: Behaviour of the focus stacking method when a crack occurs

The stack is acquired revolution by revolution, i.e. the Z-stage is moved after every rotation while the radius remains constant until the whole stack for a certain radius is acquired. Afterwards the scan moves to the next radius. As only the Z-stage is moved during the acquisition of the focus stack, the chance of alignment errors in the X-Y-plane of the image is minimized. This eliminates the need for a registration of the images in the image fusion software.

6.1. Scanning Range

In order to set the scanning range the maximum and minimum height of the disc have to be determined. As a focus stack is acquired for every radius separately, it is only necessary to determine the maximum and minimum height for the next radius and not the global ones.

The height profile of the disc is scanned with the Keyence displacement sensor. The sensor is

positioned so that it is always one step in radius ahead of the camera. Thus the maximum and minimum height for the next step can be determined during the acquisition of the focus stack. This way no extra revolutions are necessary, except for the very first stack, where one pre-scaning round is necessary to determine the range before the scan starts.

If the laser beam of the Keyence sensor is scattered, e.g. through a crack or particles on the disc surface, the sensor output value is set to the minimal value of -0xFFFFFF. These outliers need to be filtered out of the measurement data, otherwise the scanning range can not be determined correctly. In the current system this is done by simply clipping all values above or below a ±5 mm range. For the acquisition of the height profile the same filter is used and the clipped values are replaced with the last valid measurement.

6.2. Disc Warpage

The warpage of the lacquer discs is expected to be in the range of several 100 µm, because of the rigid base material. Consequently the new acquisition system is designed without any tracking of the disc surface. This has the advantage of simplifying the acquisition process.

However, if the height of the cracks is much smaller than the warpage this method is inefficient, as the stack will be a lot bigger than actually necessary to get the cracks in focus. If that should be the case for most discs, the acquisition could be extended with a method to adapt the camera trajectory to the overall shape of the disc, e.g. by using a polynomial fit of the height profile to control the Z-axis.

6.3. Implementation

The new acquisition control software could be derived from the existing one. The main differences are that no tracking of the disc surface is used and several revolutions at the same radius have to be done to acquire the focus stack. Because there are no changes in the way how the hardware is operated many of the existing VIs could be reused, which makes the implementation rather easy.

TOOD: mention interface changes

6.4. Summary

The new acquisition system acquires a stack of images with a constant spacing for every scanning radius from the minimum to the maximum height of the disc. The scanning range is determined before the acquisition with the Keyence laser displacement sensor.

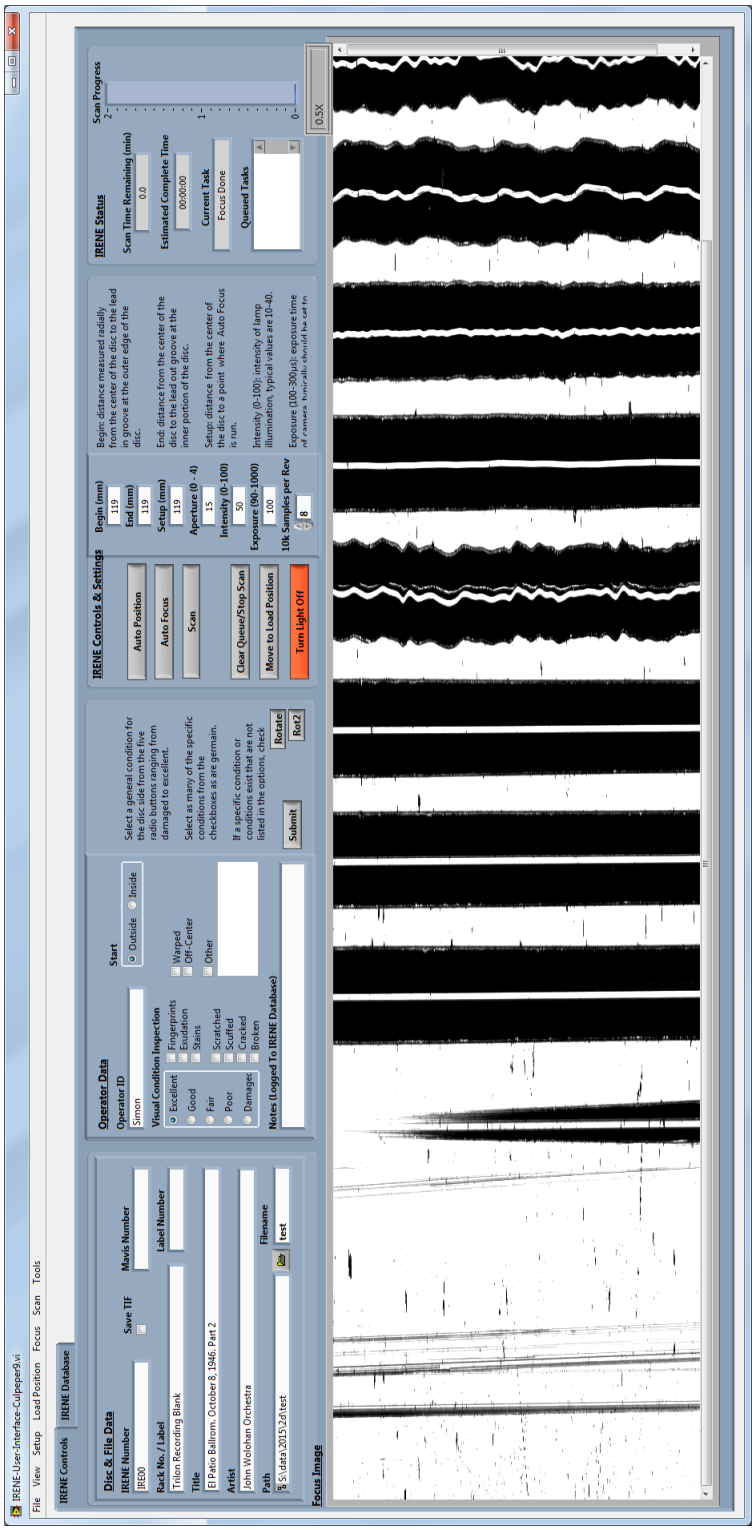


Figure 6.5.: Screenshot of the new IRENE 2D user interface

Part III.

Fusion of Images with Multiple Depths of Field

7. Related Work

Using a stack of images to enhance the depth of field of a microscope with image processing methods is not a new idea. The method was first described in 198 by R. J. Wall et al. [45]. Valdecasas et al. gives an overview over the variety of approaches that have been proposed in the literature since then [46]. The basic approach is always the same. In the first step a likelihood for being in focus is calculated for each pixel in every image in the stack. The final image is then assembled by taken the pixel with the highest measure from the stack.

Based on the nature of the likelihood function the fusion methods can be divided into three categories, pixel based, region based and frequency domain based algorithms. In the pixel based approaches the decision which pixel is used for the final image is made individually for each pixel. In the region based approaches the decision is made on the level of subparts of the image instead. Most focus measures utilize the fact that the focus of an image can be identified by looking at the high frequency content. If the focus is lost small features will become invisible due to the defocus blur, which is equivalent to a loss of high frequency information in the image. That means that the image or a region with the most high frequency content is likely the one with best focus. Pixel and region based approaches use spatial filters to identify the high frequency content. The frequency domain methods use filters in the Fourier or wavelet domain instead.

The following sections provide an overview over different approaches. Because of the vast number of publications on the topic, the literature research can not be seen as extensive in any way. Thus the cited papers are chosen as examples for different approaches and may not necessarily reflect the state of the art.

7.1. Pixel Based Methods

R. J. Pieper and A. Korpel describe basic algorithms for pixel based fusion [47]. One of the approaches is to use the pixels with the maximal or minimal intensity from the stack. The assumption herein is that the gray level of the pixel tends to get closer to the average intensity when it goes out of focus. The idea becomes clear when one thinks about a black dot on a white background. If the dot gets out of focus it will become more and more gray, as it blends together with the background because of the defocus blur. To determine if the maximum or minimum has to be used for the final pixel, the decision is made based on the difference between the average and the extrema. If the stack of images is relatively large the average intensity $\overline{I(x,y)}$ will be close to the defocused value. Therefore the value as far apart as possible from the average should be closest to the in focus value. The function $Q(x,y)$ in equation 7.1 is used to decide if the minimum or maximum should be used. If $Q > 0$ the maximum is used, otherwise the minimum.

$$Q(x,y) = |I_{max}(x,y) - \overline{I(x,y)}| - |I_{min}(x,y) - \overline{I(x,y)}| \tag{7.1}$$

Another approach is to use a difference operator similar to the Brenner or Sobel focus measure in section 5.4.3. In this case the nondirectional difference operator in equation 7.2 is used. After

the calculation of the measure, the pixel with the highest measure is then chosen from the stack. This method becomes very sensitive to noise in regions that do not contain high frequency information like plain areas. To compensate for that, the average of all pixels in the stack is used, if the derivate is below a certain threshold.

$$\begin{aligned} f_{Pieper}(x,y) = |g(x-1,y+1)-g(x+1,y-1)| + |g(x+1,y+1)-g(x-1,y-1)| \\ +|g(x,y+1)-g(x,y-1)| + |g(x-1,y)-g(x+1,y)| \end{aligned} \tag{7.2}$$

H. A. Eltoukhy and S. Kavusi propose a more sophisticated pixel based fusion approach that is tuned towards computational efficiency [48]. They use the focus measure described in equation 7.3. This method deals with plain areas by averaging the focus measure over a N x N neighborhood, to incorporate the surrounding pixels in the focus decision (see equation 7.4). This processing step can be implemented simply by convolution of the focus measure image with a N x N kernel with all values set to 1. For the decision which pixel should be used a sigmoid function is then applied to the focus measure (see equation 7.5). The proposed algorithm is aimed at situations were only two images $g_i(x,y)$ and $g_j(x,y)$ need to be fused. Therefore the final gray value can be calculated as linear combination of the two pixel values as in equation 7.6.

$$f_{Eltoukhy}(x,y) = |g(x,y)-g(x-1,y)| + |g(x,y)-g(x,y-1)| \tag{7.3}$$

$$\tilde{f}_{Eltoukhy}(x,y) = \sum_{y=-M/2}^{M/2} \sum_{x=-N/2}^{N/2} f_{Eltoukhy}(x,y) \tag{7.4}$$

$$\hat{f}_{Eltoukhy}(x,y) = \frac{1}{1+e^{-\beta \tilde{f}_{Eltoukhy}(x,y)}} \tag{7.5}$$

$$g_{fused}(x,y) = \hat{f}_{Eltoukhy}(x,y) g_i(x,y) + (1-\hat{f}_{Eltoukhy}) g_j(x,y) \tag{7.6}$$

7.2. Region Based Methods

H.-S. Wu, T. Kwok and Y. Wang introduce an approach that instead of evaluating the focus for every single pixel uses rectangular blocks [49]. The subimages are high-pass filtered by applying a gaussian low-pass filter and subtracting the output from the original image. Then the energy of the block is calculated through squaring and summation. Finally the block with the highest energy is chosen for the final image.

S. Li, J. T. Kwok and Y. Wang present a similar approach but use the spacial frequency as defined in equation 7.9 as focus measure [50]. The method is again aimed at situations were only two images are merged. After the calculation of the focus measure the decision which blocks should be copied to the final image is made. If the difference between the measure of two block exceeds a certain threshold, the block with the higher image is used. Otherwise the average of two blocks is copied to the final image. The purpose of the threshold is to make the method robust against noise in areas with low high frequency content. To further increase the robustness a majority filter with a window size of 3 x 3 pixels is used. Majority filtering means, that every block is set according to the majority of the 8 surrounding blocks. If for example

block is set to image 1 but the majority of the surrounding blocks are set to image 2, then the block will be set to image 2 as well.

$$f_{row} = \sqrt{\frac{1}{MN}\sum_{x=1}^{N}\sum_{y=2}^{M}(g(x,y)-g(x,y-1))} \tag{7.7}$$

$$f_{column} = \sqrt{\frac{1}{MN}\sum_{y=1}^{M}\sum_{x=2}^{N}(g(x,y)-g(x-1,y))} \tag{7.8}$$

$$f_{spatial} = \sqrt{f_{row}^2 + f_{column}^2} \tag{7.9}$$

W. Huang and Z. Jing present an evaluation of different focus measures for block based image fusion methods [51]. The compared measures are variance, Tenengrad, Laplacian (for all three see section 5.4.3), spatial frequency and sum modified Laplacian (see equation 7.11). The sum modified Laplacian focus measure uses the absolute values of difference to avoid cancellation of positive and negative values (see equation 7.10). T is a threshold value similar to the one used in Tenengrad. In the tests only two source images are used. The blocks are only chosen based on the focus measure values, no additional filtering is applied. The bottom line of the evaluation is that the sum modified Laplacian provides the best quality of the output images.

$$ml(x,y) = |2g(x,y)-g(x-1,y)-g(x+1,y)| + |2g(x,y)-g(x,y-1)-g(x,y+1)| \tag{7.10}$$

$$f_{SML} = \sum_{x=1}^{N}\sum_{y=1}^{M} ml(x,y) \text{ for } ml \geq T \tag{7.11}$$

The regions used for the evaluation of focus measures do not necessarily be simple geometrical shapes. S. Li and B. Yang present an approach that is based on segmentation of the image in regions with high similarity [52]. In the first step the average of the to fused images is used to construct regions using the normalized cuts algorithm. After the construction of the regions the focus for each of them is determined by calculation of the spatial frequency. The regions with the highest spatial frequency are then copied to the fused image.

7.3. Frequency Domain Methods

As already mentioned the focus of the image can be determined by looking at the high frequency content of the image. Consequently one could use a Fourier transform and look at the spectrum of the image to determine it's focus. However, the Fourier transform does not provide any information were the sharp feature in the image reside, which makes it unsuitable for image fusion. This problem can be solved by using a Wavelet transform. The Discrete Wavelet Transform (DWT) is one of the standard methods to fuse images from different sensors and it can also be used to fuse focus stacked images. DWT was applied to multi-focus image fusion for the first time by H. Li, B. S. Manjunath and S. K. Mitra [53].

Like the Fourier Transform, the wavelet transform is a complex mathematical tool and giving a detailed description is out of scope of this thesis. Introductions to the topic can be found for example in the tutorials by R. Polikar [54] and C. Valens [55]. The basic idea of the Wavelet

transform is to convolute the signal with brief oscillating functions instead of infinite periodical ones, as in the Fourier transform. In the continuous case the Wavelet Transform can be described as the convolution of the signal with scaled and dilated versions of the Wavelet function $\psi(\tau, s)$ (see equation 7.12).

$$\Psi_x^t(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t)\psi^* \left(\frac{t-\tau}{s} \right) \tag{7.12}$$

In signal processing applications the DWT is often implemented as filter bank. The discrete signal $s[x]$ is filtered with a low-pass filter $w_{low}[x]$ and high-pass filter $w_{high}[x]$ like in equations 7.13 and 7.14.

$$t_{high}[k] = \sum_x s[x] w_{high}[2k - x] \tag{7.13}$$

$$t_{low}[k] = \sum_x s[x] w_{low}[2k - x] \tag{7.14}$$

The output of the lowpass filter is then downsampled and feed again into the same filter bank. This process can be repeated until the lowest desired resolution is reached. This allows a multi-resolution analysis of the image using a minimal number of filters. In image processing applications the filters are first applied to the columns then to the rows, which creates the directional high-high $h_d(x, y)$, high-low $h_{v(x,y)}$, low-high $h_h(x, y)$, and low-low $h(x, y)$ outputs (see figure 7.1).

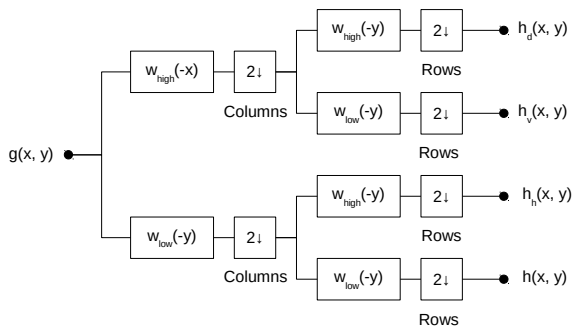
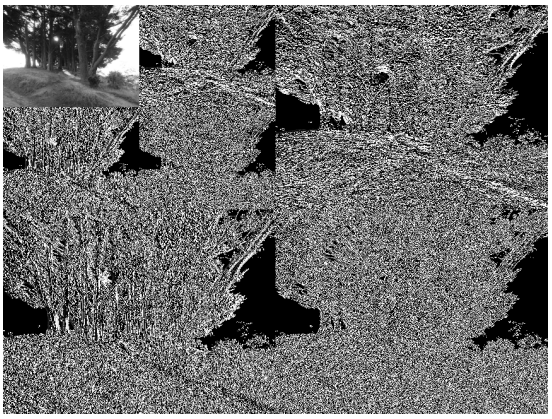


Figure 7.1.: Block diagram of a filter bank for 2D wavelet transform

The DWT based image fusion algorithm proposed in [53] works as follows. In the first step the DWT is calculated for both images. In the second step the larger absolute values from both sets of coefficients are chosen. From this fused coefficients the inverse DWT is then calculated. Various improvements of this approach have since been proposed, see for example [56, 57].

(a) Input image



(b) Output coefficients

Figure 7.2.: Example of a 2 stage discrete wavelet transform

8. Design of the Image Fusion Algorithm

For the evaluation of several approaches for the fusion algorithm a simple GUI is implemented in C#. The idea is to evaluate several algorithms on a small test image without having to deal with 300 MB to 800 MB images. As test image a section of a lacquer disc recorded in 1946 by the John Wolohan Orchestra (further referred as John Wolohan disc) is chosen, because it shows the typical patter of deterioration of the lacquer coating (see figure 8.1).

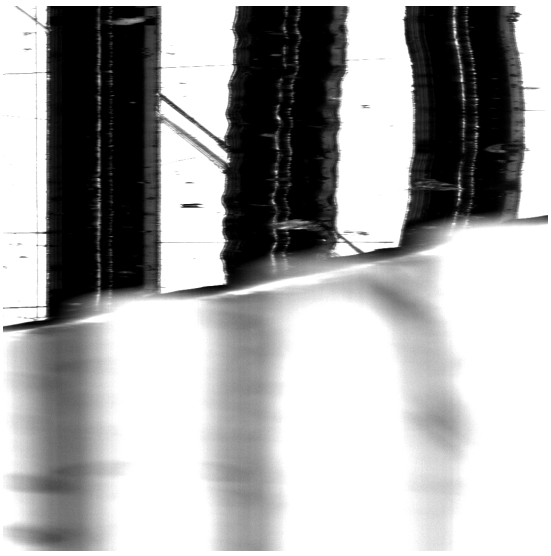
Figure 8.1.: Recording of the John Wolohan Orchestra made on October 8, 1946

8.1. Evaluation of Fusion Algorithms

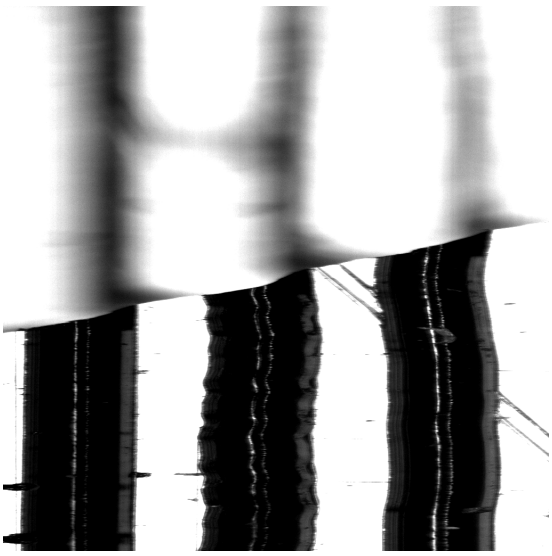
A section of 4096 x 8192 pixels from the scan is used as test image. This section of the scan is chosen because it shows a very large height variation of about 1000 µm. That means that in order to create a fused image with good focus regions of all 33 image in the stack need to be fused. This gets especially clear when one looks at images from the bottom and the top of the stack 8.2.

One problem in the design of the image fusion algorithm is, that the images can not be just compared against a target image. Obviously the output should contain the in focus regions from every image, but how the border regions should look like is not clear. One method is to compare with the image that shows the either side of the crack at a time, but this is cumbersome and does not provide an overall impression of how a well merged image should look like. Therefore a small part of the test image was merged using the GIMP image manipulation software to create

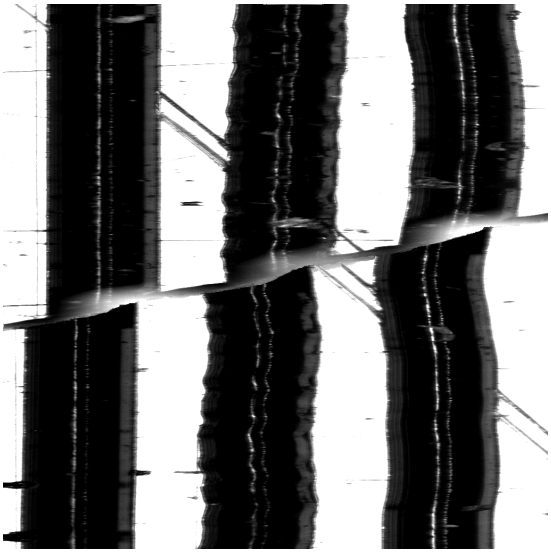
a mock-up. Figure 8.2c shows the mock-up which consist the images 8 of 33 at the top and 32 of 33 at the bottom of the crack.



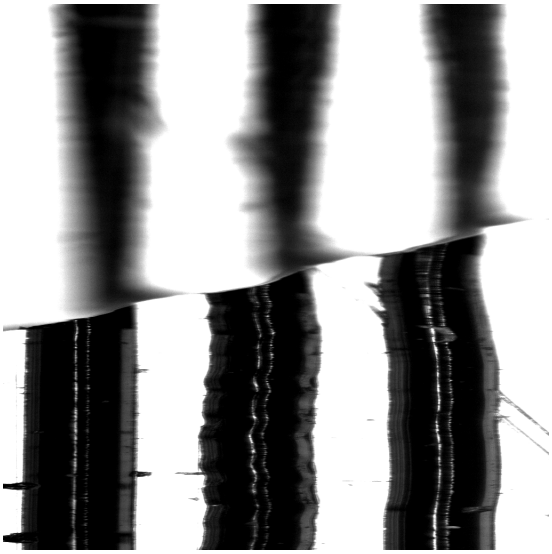
(a) Test image 8/33 from the stack



(b) Test image 23/33 from the stack



(c) Images 8/33 and 32/33 merged using GIMP



(d) Test image section scanned with the tracking acquisition method

Figure 8.2.: Test images

The approach for the evaluation of fusion algorithms is to start with simple algorithms and the proceed to more sophisticated ones until a satisfactory solution is found. This way chances are high that a solution with low complexity can be found that then can be optimized to work with the large images. To ensure comparability of the output of the different algorithms the Brenner focus measure, respectively the squared horizontal first order difference was used in all cases.

The first step of an image fusion algorithm is usually a registrations of the images. In this case this can be omitted because the acquisition produces well aligned images (see chapter **??**).

8.1.1. Average Image

The simplest approaches for image fusion is to create an average image by adding all images of the stack. This approach was successfully used to fuse small parts of two scans by Fischer [7]. Although his work shows that the result quality of the fused image decreases rapidly if the number of images is increased. This method is suitable to get an overview over the content of the whole stack but all sharp edges get blurred through the averaging process.

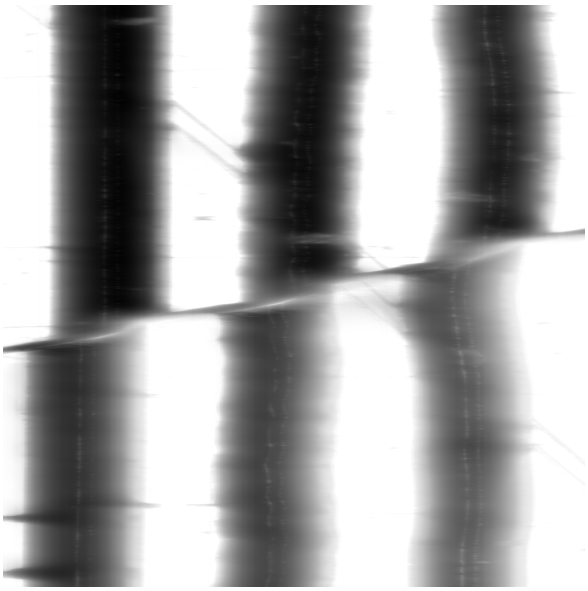
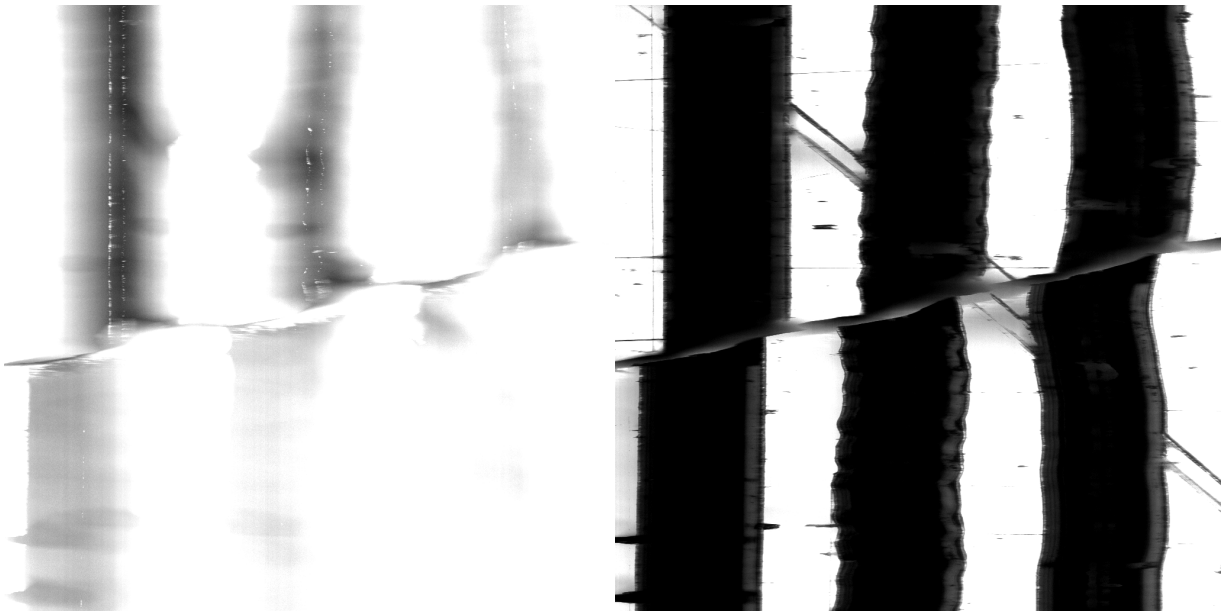


Figure 8.3.: Detail of the average image

8.1.2. Minimum and Maximum Image

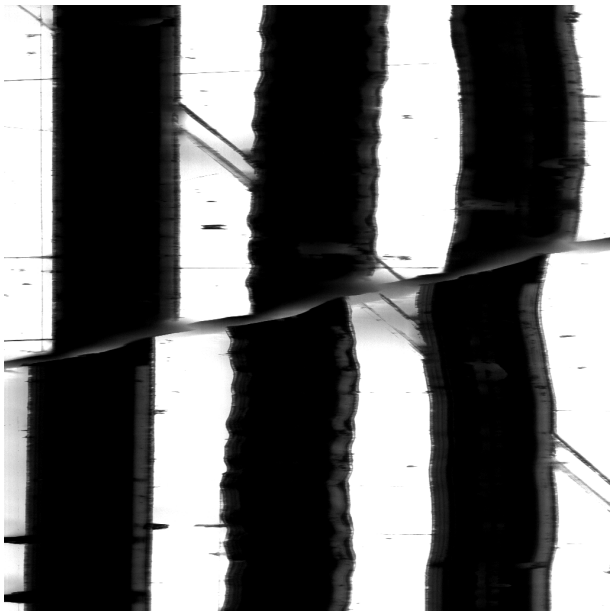
The groove bottom appears normally as white line when in focus and grays out when the focus is lost. Therefore it should be possible to extract the groove bottom by searching the maximum of each pixel in the image. In contrast the sides of the groove appear black when in focus and become narrower when the focus is lost. This thinning out of the groove sides is caused by the high camera gain that is normally used for the acquisition. When the focus is lost the edge of the groove side get blurred and appear therefore in a gray. Through the high gain these gray areas are saturated into white, what makes the groove edges appear narrower. Because the groove sides appear the widest in the in the image with best focus, the groove sides should appear in focus in the minimum image. If it is possible to extract the the groove bottom separately in the minimum and maximum image, both images could be fused with a method similar to the decision function propsed by [47]

Figure reffig:MaxImage shows the minimum and maximum image computed from the test stack. The groove sides are extracted nicely in the maximum image. Unfortunately the groove bottom is not well preserved in the maximum image, mainly because the groove sides are completely blurred. Hence the contrast between groove side and bottom is lost. The same happens in the image fused using the decision function propossed in [**?**] (see also section 7.1). The groove sides are preserved well but the groove bottom is only vaguely visible. In sum this method does not look very promising and is hence not investigated further.



(a) Maximum Image

(b) Minimum Image



(c) Image fused using Pieper decision function

Figure 8.4.: Minimum and maximum based approaches

8.1.3. Region Based Fusion

Another simple approach to fuse images is to divide the image into regions and chose the image with best focus for each region. This can be done by calculating a focus measure for each region and then choosing the region with the highest measure from the stack.

Tile Fusion

A straightforward approach for partitioning the image is to divide it into rectangular sub images, in the following called tiles. To fuse the images the focus measure for all tiles for each image in the stack have to be computed. Then for each region the tile with the highest measure needs to be identified. The corresponding index is then saved in an array, called fusion map. Finally the tiles are copied into the fused image according to the fusion map. Similar approaches have been proposed in [49] and [50] (see section 7.2).

This approach works relatively well if the size of the rectangles is chosen bigger than the width of the groove. The example in figure 8.5a shows the result of a tile fusion with a tile size of 256 x 256 pixel. The image has a good focus overall, but the borders between the tiles are visible at some places. These discontinuities could lead to artifacts in the audio.

If the tile size is lowered to 64 x 64 pixel some of the tiles do not contain edges any more. In this cases the focus measure becomes zero and the decision which tile should be taken is ambiguous. The algorithm was designed to fall back to the lowest tile from the stack, which causes the white holes in the image (see figure 8.5b). This effect can be reduced by smoothing the the fusion map with a median filter. In this case a kernel with a size of 9 x 9 pixel was used for the smoothing. The smoothing decreases the discontinuities at the border of tiles as well (see figure 8.5c). The downside of the smoothing is that the focus reduced at steep changes in the disk surface (see figure 8.5d).

Line Fusion

As the image is acquired in lines of 4096 x 1 pixel to use this line as regions is an natural choice. In this approach the Brenner measure is calculated for every line for all images in the stack. Then for every line the one with the highest measure is copied into the fused image.

For most parts of the image this approach works very well, but it does not work near cracks. If the crack is not perfectly radial either the line corresponding to one of the sides of the cracks is chosen, depending which one occupies the bigger part of the line 8.6. Another problem could be cases where the disc has a slope in the radial direction, so that the line in one image does not provide could focus over the whole length. If this method is chosen, a work around for these regions has to be found.

8.1.4. Pixel Based Fusion

Pixel Fusion

The contrary approach to using regions is to determine the focus for every single pixel. The approach is basically the same as the tile based fusion but the tiles now have the size of a single pixel. As focus measure for every single pixel the squared first order difference is calculated. Then for every pixel the one with the highest measure is copied into the fused image. This

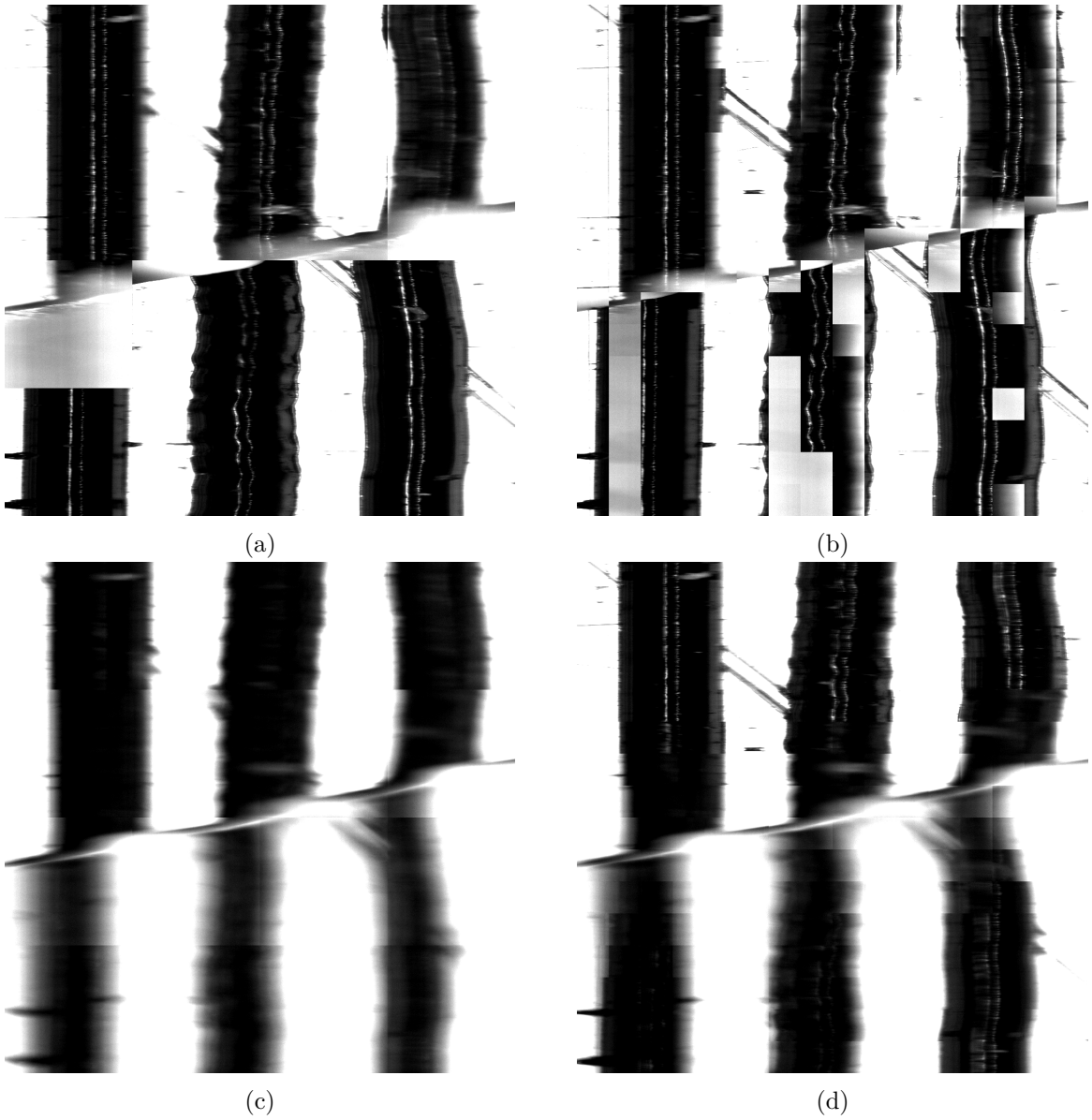


Figure 8.5.: Tile fused images ((a) tile size: 256 x 256 pixel, (b) tile size: 64 x 64 Pixel, (c) tile size: 256 x 256 pixel, smoothing kernel size: 9 x 9 pixel, (d) tile size: 64 x 64 Pixel, smoothing kernel size: 9 x 9 pixel

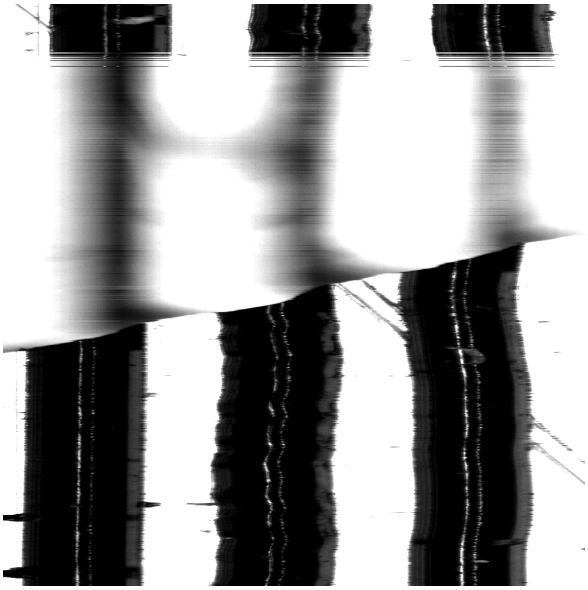
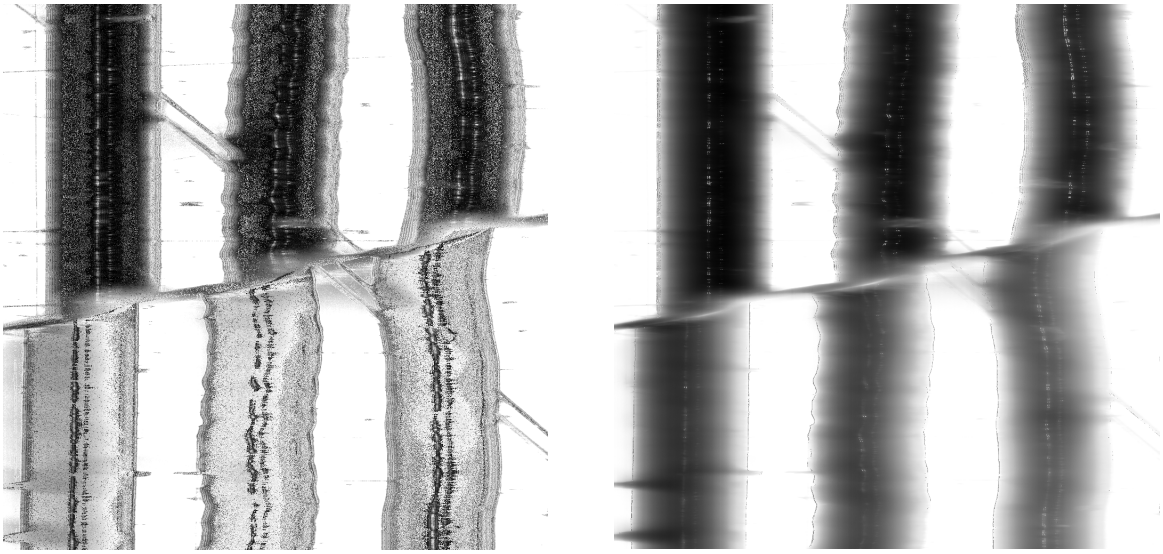


Figure 8.6.: Detail of the line fused image

approach obviously fails for even regions, which leads to the high amount of white pixels. At this points the focus measure for all pixels in the stack is zero, therefore the lowest pixel is taken from the stack (see figure 8.7a).

This method can be somewhat improved by applying a threshold to the focus measure, similar to the method proposed in [**?**]. If the maximal focus measure for a given pixel is below the threshold the average of the pixels of the stack is used (see figure 8.7b). In both cases the algorithm fails at the plain areas in the image.

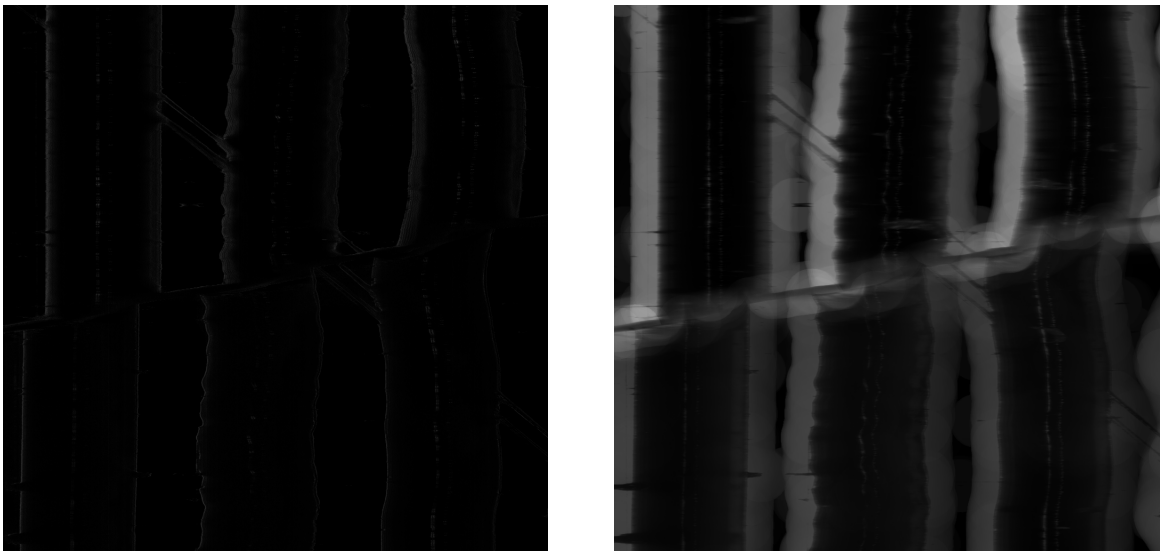


(a) No thresholding (b) Threshold = 80

Figure 8.7.: Pixel fused images

Mask Fusion

Instead of determining the pixel with the highest measure fore every pixel an alternative approach is to use directly the derivative image as mask for the fusion. The idea is to multiply every image with its derivative and the add all images. If the pixel has a high derivative the multiplication increases the value, a low derivative will decrease the pixel value. This should boost the pixel with good focus. The distribution of the values can afterwards be corrected again through normalization.



(a) (b)

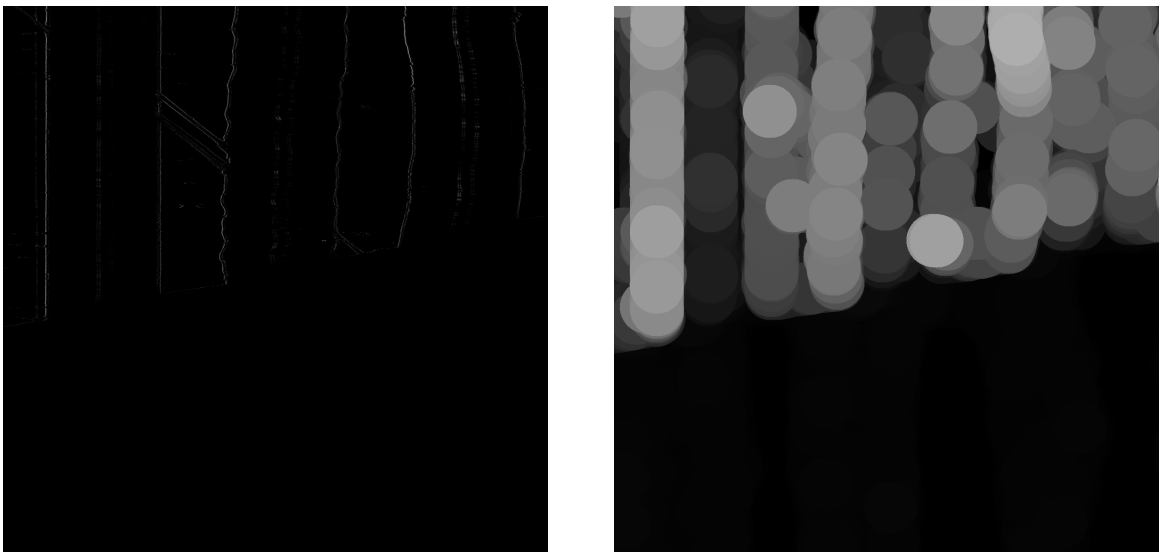
Figure 8.8.: Map fused images ((a) without dilation of the derivative image, (b) with dilation of the derivative image, structuring element size: 101 x 101 pixel)

Obviously the even regions are completely lost in this approach, only region with edges remain in the image (see figure 8.11. Therefore a morphological dilation is applied to the derivative image to spread out the high derivative points over even regions. The dilation basically works as maximum filter, were every pixel is set to the maximum of it's neighbors [41]. Which neighbors are evaluated for the determination of the maximum, can be controlled by the choice of the structuring element. In common use are rectangular, elliptical and cross shaped structuring elements. In this case a rectangular structuring element with a size of 101 x 101 pixel was used.

In this approach there is no need to identify the pixel with the best focus. But this comes with the drawback that for every pixel information from all pixel in the stack is mixed in. As the depth of field of the system is very narrow in comparison to the height variation of the disc, it can be assumed that most regions can be covered with good focus when only one or maybe two images from the stack are considered. From that point of view this method is an overkill.

8.1.5. Map Based Fusion

The idea of this approach is to combine the strengths of the previous ones. It should use regions to preserve even areas but this regions should also adapt to discontinuities in the disc surface like cracks. Theoretically one could use region segmentation like for example the watershed algorithm to detect the area between cracks and use these as regions for the image fusion. But



(a) (b)

Figure 8.9.: Derivative images ((a) squared horizontal derivative, (b) derivative after dilation)

this regions would not consider the loss of focus through the unevenness of the surface caused by the disk warpage and bulges of the delaminating lacquer. Therefore an approach that is base on constructing regions from the focus information is chosen.

The basic approach is to calculate the focus measure for each image and then construct a map, which shows for every pixel the index of the image with best focus. The advantage of using such a fusion map is that it can be further processed with common image processing methods to achieve the best compromise between good focus and a smooth fusion.

The processing steps are as follows. At first the focus measure for each image is computed, here again the horizontal derivative is used for the first evaluation. Afterwards a morphological dilation is applied to construct focus regions. That means that the narrow lines from the edge detection are grown with a dilation until both top edges of the groove blend together and the whole groove area becomes evenly colored. For the dilation a circular structuring element with 21 pixel diameter is used and three rounds of dilation are applied. A circular structuring element is preferred to a rectangular one, as it introduces less artifacts in the fused image. From the focus measure images the fusion map can be constructed. It basically contains the index of the image with best focus for every pixel (see figure 8.10). To get smooth transitions at regions with abrupt height changes on the surface a median filter is then applied to the fusion map. For the median filter a rectangular kernel of 101 x 101 pixel is used. The result of the fusion process can be seen in figure 8.11.

This method yields the best result so far, but still creates a lot of visible artifacts in the fused image. Furthermore the dilation and the smoothing have a long computation time because of the big kernel sizes.

8.1.6. Conclusion

After several approaches have been evaluated the most promising approach for the further development has to chosen. This way the further improvements can be focused one a single approach.

Figure 8.10.: Fusion map

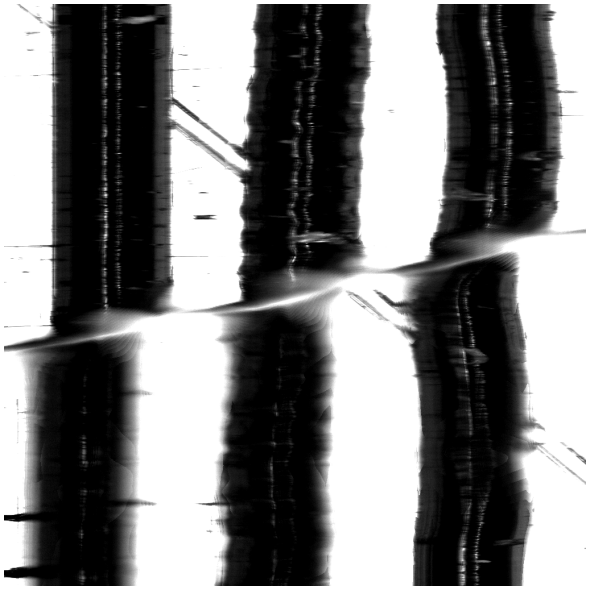


Figure 8.11.: Map fused image

All simplest approaches, averaging, minimum and maximum image lead to mediocre results and are therefore not considered for further developments. The line based fusion approach is relatively efficient but cracks in the image are not handled well. To make this approach usable a method to identify and fill in the gaps with needs to be found.

The tile based approaches stand out because bigger features are well preserved in the fusion process. To overcome the issue of visible borders between the tiles a suitable smoothing method has to be found. One approach could be to use overlapping tiles and to apply a windowing function before the overlapping areas are added together. Similar approaches are often used in image stitching processes [**?**]. Another problem is that the tiles need to be sufficiently small so that the areas near cracks have a good focus. As already mentioned in section 8.1.3 reducing the size of the tiles below the groove width can lead to problems. Therefore the handling of even areas would need to be improved as well.

The fusion map based approach provides the best results so far. It also maintains the focus best near cracks in the disc. The drawbacks are that the morphological dilation creates artifacts in the image and that the process is computationally expensive. The efficiency of the algorithm could be improved by creating the fusion map from downsampled versions of the images. Through downsampling not only the number of pixels to processed would be reduced also smaller kernels could be used.

The squared horizontal derivative that is used for edge detection in the fusion map approach produces very narrow lines. An edge detection algorithm that produces wider edges could possibly reduce the amount of morphological dilation operations necessary. This could further reduce the computation time of the fusion process.

Another aspect that needs further investigation is the narrowing of the groove edges when the focus is lost, mentioned in Section 8.1.2. The high gain setting used in the acquisition cause blurred regions to be saturated to white, therefore the groove appears narrower if the focus is lost. To reduce this effect could facilitate the fusion process. Therefore the possibility of scanning with lower gain setting should be investigated.

8.2. Improvements of the Map Fusion

8.2.1. Reducing the Camera Gain

The digital camera gain is a multiplication factor that is applied to all pixel values after analog to digital conversion. The black level is an offset that is added to the pixel values after the multiplication [27]. Both operations are done in saturation arithmetic, i.e. if a value exceeds the maximum it is clamped to the maximum and if it exceeds the minimum it is clamped to the minimum. Figure 8.12 shows the relation between the sensor value, gain, black level and gray value. Normally these transformations are used to make optimal use of the 8 Bit integer range of the intensity value, which means that the darkest pixels in the image are 0 and the brightest are 255.

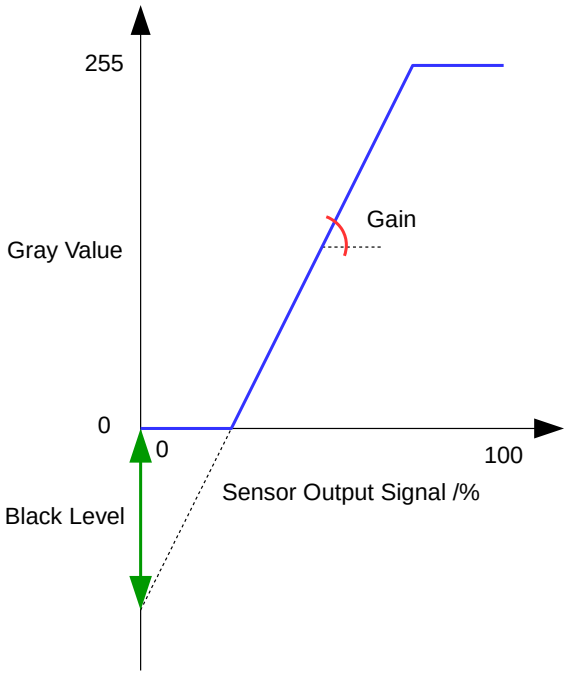


Figure 8.12.: Camera gain and black level

In this case gain and black level are usually set so that the saturation occurs on purpose. With a small gain setting the scratches and other features on the disc surface would be visible in light gray. These features could introduce errors in the detection of the groove edge. Therefore the gain is set so high that these features are saturated to white. As side effect the large gain also lets the groove sides appear in gray. To compensate the large gain and increase the perceived contrast of the image again the black level is set to a large negative value. This creates the typical images with the gaps between the grooves and the groove bottom in white and the groove sides in dark black.

The downside of the high gain setting is that it causes blurred regions in out of focus images to disappear. The blurred regions of the groove in light gray are saturated to white, which creates an artificial edge in the image. These edges can not be easily distinguished from the actual groove edges, hence they complicate the fusion process. As both gain and black level are applied in the digital domain, the could as well be applied after fusion of the images.

Several scans of the same section of the John Wolohan disc with different setting were made,

to determine if scanning with lower camera gain is feasible. Details of the scans can be seen in figure 8.13. If the gain is lowered to 1024 the groove edges appear blurred when out of foucs, as expected. The problem is that with the lower gain the groove bottom becomes so dark, that it is barely visible. This makes it hard to judge the quality of the scan by eye. That can be important for the person who operates the scanner, to adjust the lighting intensity or exposure time. A possible work around could be to scan with a lower gain but show the operator a processed image.

In conclusion the scan could be made with lower gain. However, it would be preferable to have a fusion method that works with the current settings.

8.2.2. Focus Measure

For the evaluation of the algorithms the horizontal first order difference was used as focus measure in all cases. It was chosen mainly because it is used in the Brenner focus measure and the IRENE focus measure, which both provided accurate results in the evaluation (see section 5.4.3). However in this application the focus measure is only needed two determine which image from the stack should be used for a certain pixel and not to determine the position of the focus plane accurately, like in autofocusing applications. Therefore a certain loss in precision is acceptable.

The problem of the first order difference in this application is that it provides very narrow edges. Consequently the focus region for the fusion map have to be constructed from very few points, which makes several rounds of morphological dilation necessary. With an edge detection that provides wider grooves, the number of morphological dilations could probably be reduced and the construction of the fusion map can be accelerated. Three common edge detection algorithms, Sobel, Laplacian of Gaussian and morphological gradient are compared with the first order difference, to see if they provide better results.

The Sobel operator is a widely used edge detection algorithm that basically computes the gradient of the image. The working principle is explained in section 5.4.3. The Laplacian of Gaussian operator determines the second order derivative of the image, it is explained in detail in section **??**. The morphological gradient, despite it's name, does not actually compute a gradient of the image. It is the difference between a morphological dilation and erosion of the image. The dilation expands bright regions in the image, the erosion shrinks them. The difference between them basically shows the boundaries between regions in the image.

Figure 8.14 shows the output images of the first order difference, Sobel, Laplacian of Gaussian and the morphological gradient. If one looks at the images it becomes clear that the Sobel edge detection and the morphological gradient provide much better results. The downside of these two algorithms is only that they emphasize the blurred groove at the lower part of the image as well. As these blurred parts of the groove are darker than the in focus parts, a thresholding operation was applied to see if these section can be suppressed. The thresholding sets all pixels below a certain limit to 0, all pixels with a higher gray value remain. Figure 8.15 shows the result of the thresholding operation. The thresholding can not eliminate the blurred regions after the Sobel edge detection, as there are a lot of pixels with high gray levels. On the morphological gradient the thresholding operation works well, the blurred regions disappear completely. Because of these beneficial properties, the morphological gradient is used in the further development of the image fusion algorithm.

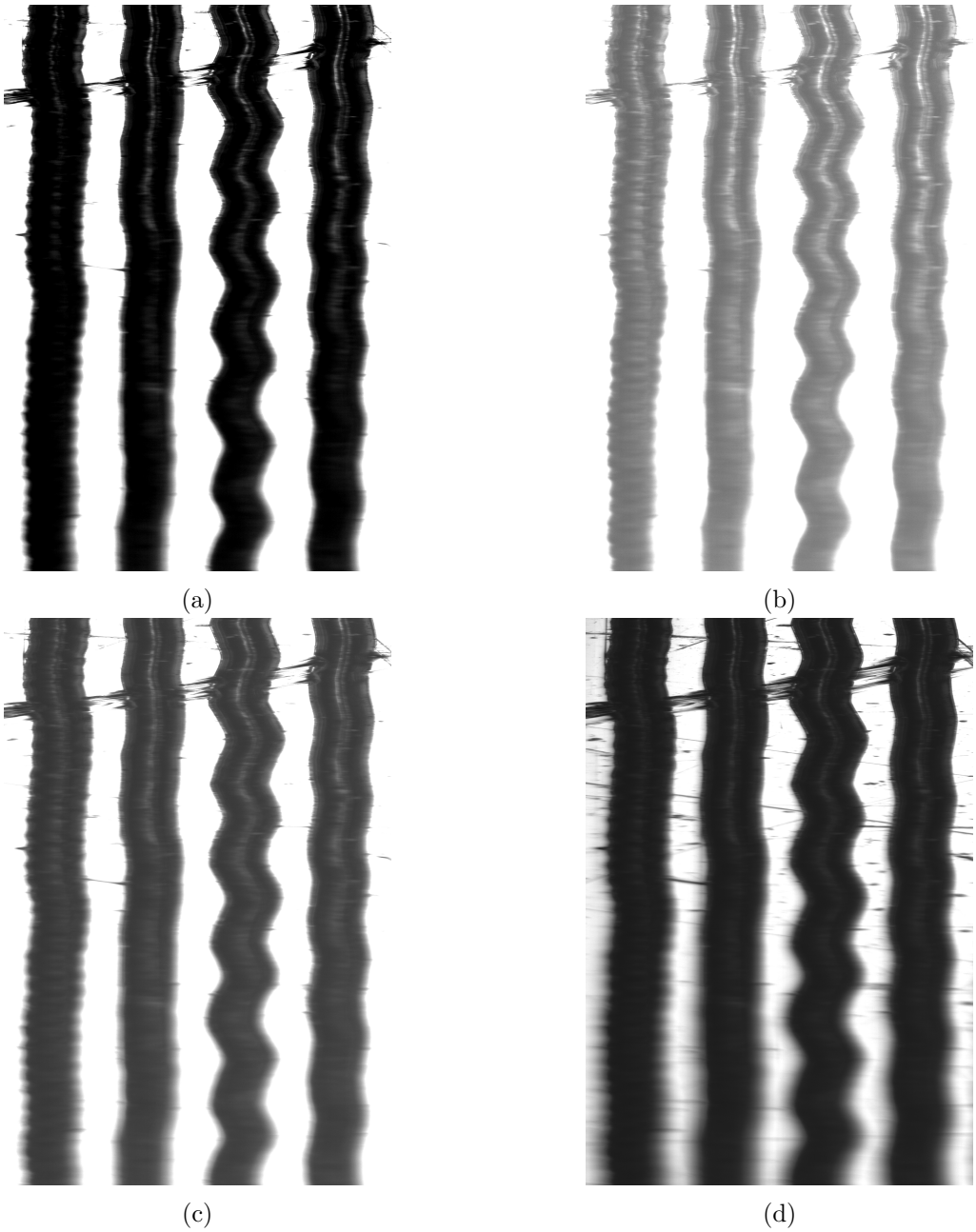


Figure 8.13.: Comparison of different camera settings ((a) gain: 2047, black level: -2048, (b) gain: 2047, black level: 0, (c) gain: 1024, black level: 0, (d) gain: 512, black level: 0

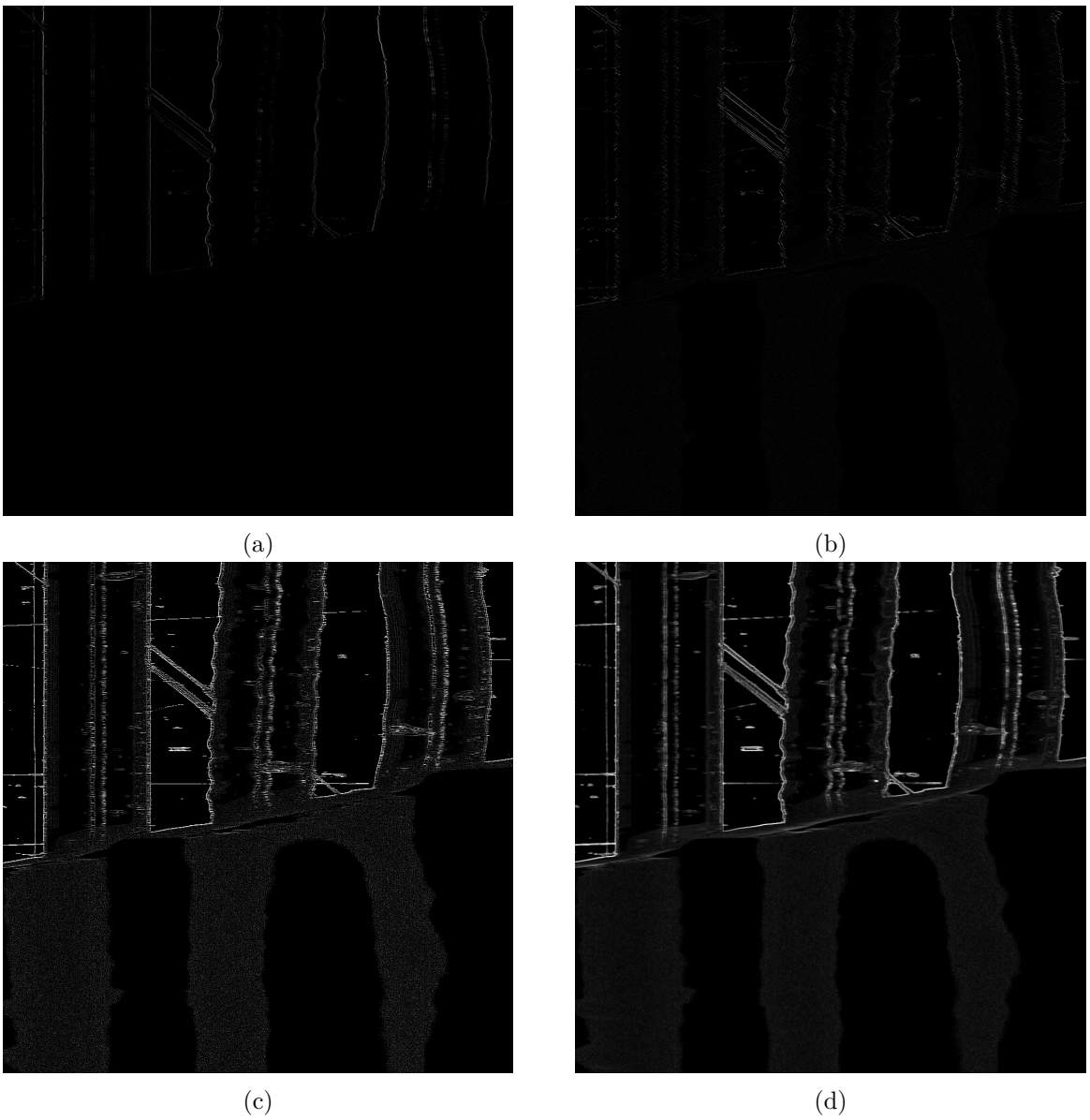


Figure 8.14.: Comparison of edge detection algorithms ((a) horizontal first order difference, (b) Sobel, (c) Laplace, (d) morphological gradient

8.2.3. Downsampling

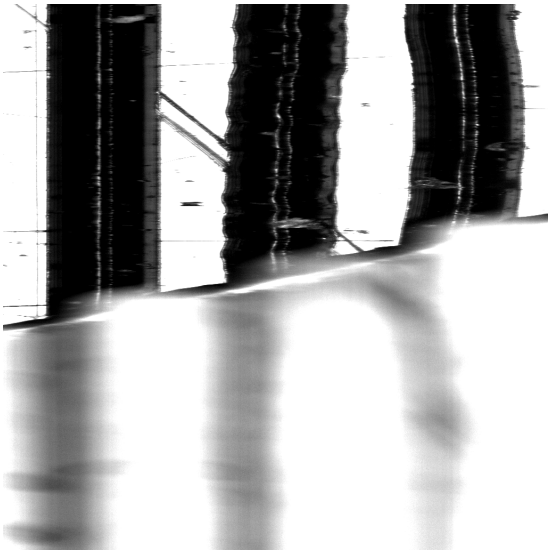
To reduce the computational workload of the image fusion the images can be downsampled to a lower resolution. The images will be downsampled after the edge detection, because the downsampling could reduce the accuracy of the edge detection.

Downsampling can be done with several interpolation methods to determine the values of the remaining pixels. The most often used interpolation methods in image processing are nearest neighbor, bilinear and bicubic interpolation [41]. Nearest neighbor simply assigns the gray value of the closest pixel in the full sized image to the pixel in the downsampled image. This approach is computationally inexpensive, but it can create artifacts in the image. The bilinear interpolation uses the four nearest neighbors to calculate an estimate of the new pixel. The bicubic interpolation uses the sixteen nearest neighbors for the estimate. As the image after the edge detection is relatively sparse, the choice of the interpolation method is critical. However, there is no need to preserve the detail of the groove. The bright pixels remaining after the downsampling will be used more like seed points for the construction of the focus regions via dilation later. Therefore a certain loss of information is acceptable. Figure 8.16 shows the same image downsampled by factor 4 with nearest neighbor, bilinear and bicubic interpolation.

With the nearest neighbor interpolation parts of the groove edge are lost resulting in a unsteady dashed line. The bilinear interpolation preserves the edges much better. From the bilinear to the bicubic interpolation the improvement is small. From this comparison the bilinear interpolation seems to be the best compromise between image quality and computational efficiency. Therefore it is the interpolation method of choice for the further development.

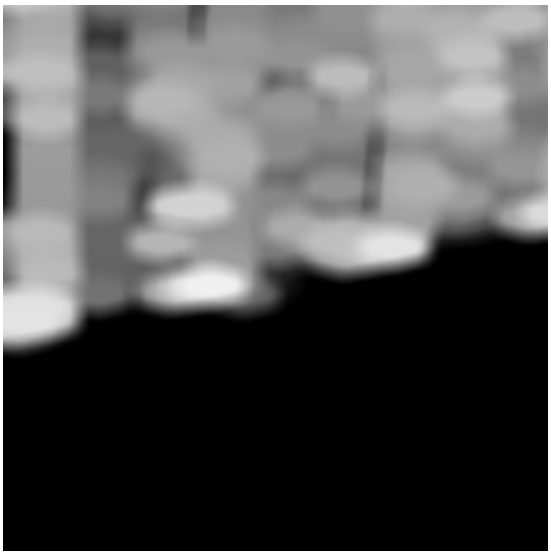
8.2.4. Construction of the Focus Map

From the downsampled edge detection images focus regions need to be constructed. From the focus regions of every image the fusion map is the compiled. There are several parameters which influence the construction of the focus region and the focus map. These factors do not only influence the quality of the output image, they also affect the efficiency of the fusion. As the fusion algorithm should work on stacks of 5 to 35 800 MB images a middle ground between image quality and computational burden has to be found. As often in image processing there is no straight forward way to determine the values for each parameter. Different values for each parameters need to be tried out, in a trial-and-error way to find a combination that produces good results. The following sections give an overview over the parameters and how they influence image quality and efficiency of the algorithm:



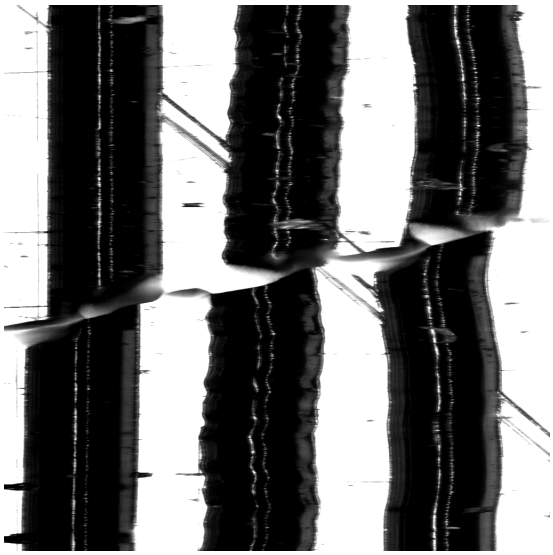
(a)

(b)



(c)

(d)



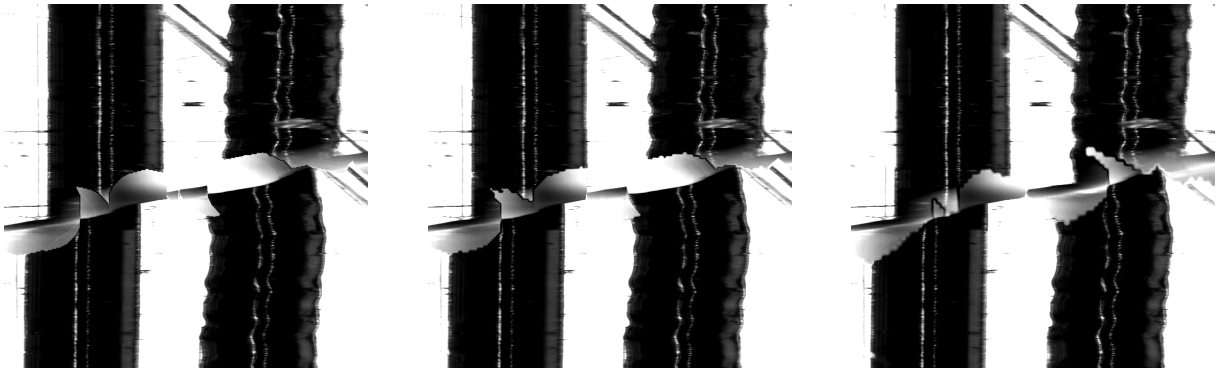
(e)

Figure 8.17.: Processing steps of the image fusion ((a) image from the stack, (b) edge detection, (c) focus regions, (d) fusion map, (e) fused image

Downsampling Rate

The downsampling rate is ratio between the size of the image and the downsampled version, e.g. a downsampling rate of 4 means that the downsampled image is 4 times smaller that the original. It sets the lower boundary for the level of detail in the focus map. On the other hand With a higher rate the images containing the focus regions and the focus map become smaller, that means the RAM usage is lower. Also can the focus regions be constructed with a smaller dilation kernel and fewer steps, i.e. the computation time is lower.

Powers of 2 between 2 and 64 have been tried as downsampling rate, the optimum seems to be around 4. As it is not possible to get good results with the same structuring element size, the size has to be adapted the downsampling rate. With a rate of 2 the dilation kernel still needs to be rather large, with a rate of 8 or higher some detail in areas around cracks get lost. Figure 8.18 shows the output for a downsampling rate of 2, 4 and 8. A round structuring element was used in all cases.

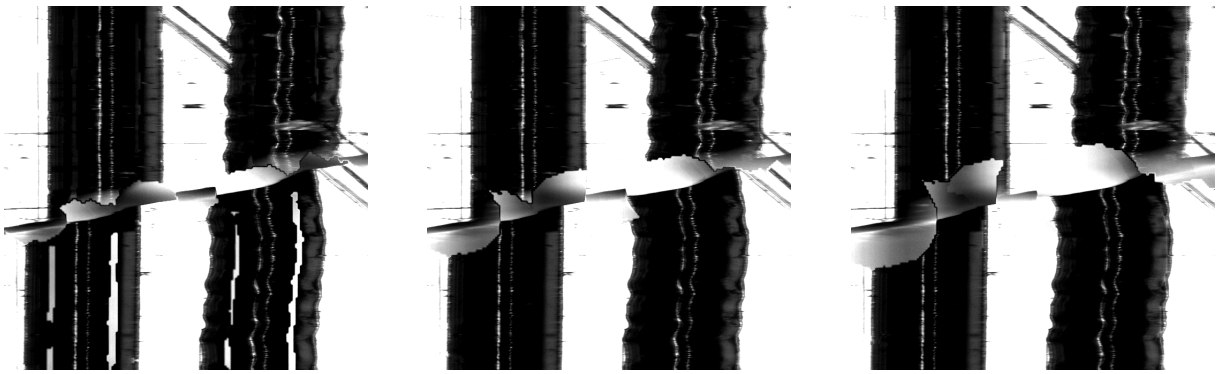


(a) (b) (c)

Figure 8.18.: Image fusion results with different downsampling rates ((a) downsampling rate: 2, structuring element diameter 21 , (b) downsampling rate: 4, structuring element diameter 11, (c) downsampling rate: 8, structuring element diameter 7

Size of the Structuring Element

The size of the structuring element influences the level of detail of the focus map as well. With a smaller element the border of the focus region will be more detailed. But with a larger element the focus region can be grown faster, in fewer dilation rounds. For a downsampling factor of 4 a structuring element of about 11 seems to work best. With smaller sizes the focus regions do not cover the whole groove area, with bigger sizes more information around cracks gets lost (see figure 8.19).

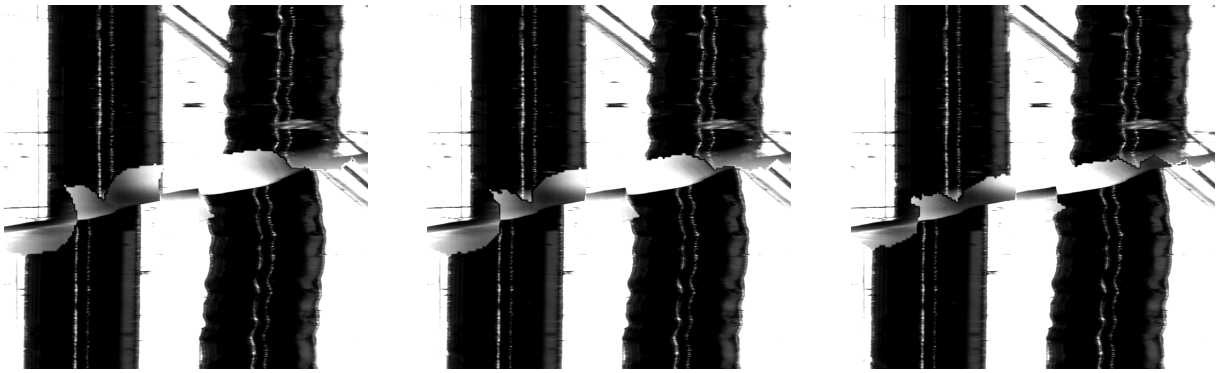


(a) (b) (c)

Figure 8.19.: Image fusion results with different sizes of the structuring element ((a) 7 x 7 pixel , (b) 11 x 11 pixel (c) 15 x 15 pixel

Number of Dilation Rounds

The dilation can be applied to the image several times. Although it is less efficient to apply several rounds with small kernel, it leads to better results regarding preservation of image details. Using a smaller kernel and several rounds produces the better results in cracked regions. Image 8.20 shows the fusion result for different number of dilation rounds with different kernel sizes. Currently 3 dilation rounds are used.

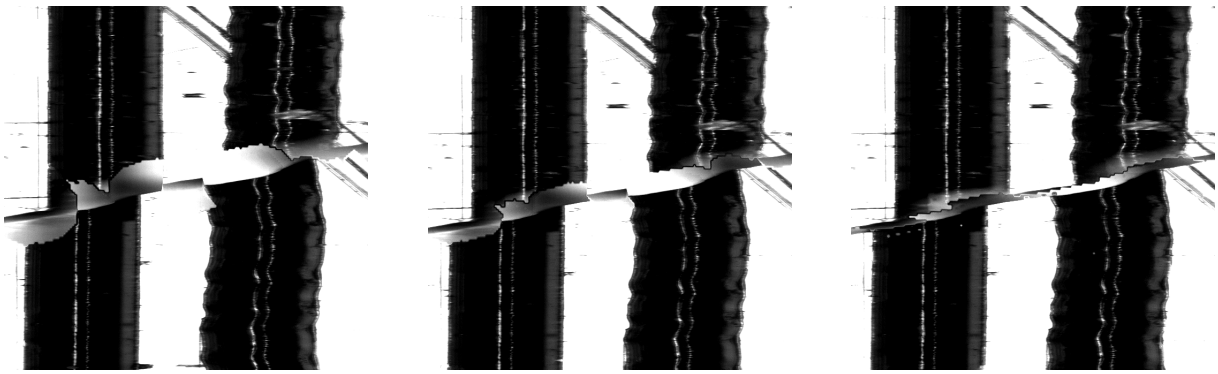


(a) (b) (c)

Figure 8.20.: Image fusion results with different number of dilation rounds ((a) 1 round, SE size 33 x 33 pixel (b) 3 rounds, SE size 11 x 11 pixel, (c) 6 round, SE size 5 x 5 pixel

Shape of the Structuring Element

The shape of the structuring element influences the shape of the focus regions. Rectangular shaped structuring elements will create straight lines at the border of the focus region, elliptical ones will create rounded borders. Rectangular structuring elements have found to introduce visible artifacts in the image, therefore elliptical ones are preferred. Elements with an aspect ratio of about 2:1 (width:height) lead to better result around cracks. The reason becomes clear when one looks at the edge detection images (see 8.14. With these kernels the focus region grows faster horizontal direction that vertically. These causes the edges of the groove to merge to one focus region, but the region does not grow over horizontal cracks. If the kernel height is set to the focus region becomes holes, which appear as small white spots in the finial image.



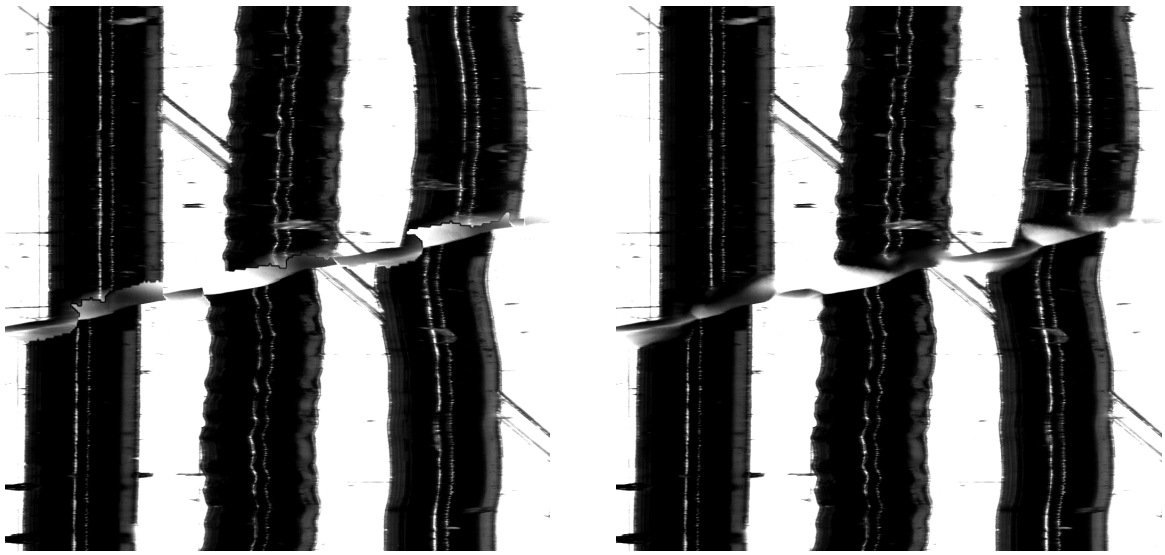
(a) (b) (c)

Figure 8.21.: Image fusion results with different aspect ratios of the structuring element ((a) 11 x 11 pixel (b) 11 x 7 pixel, (c) 11 x 3 pixel

Smoothing of the Focus Map

After the construction of the focus map, the map is smoothed to avoid artifacts in the fuse image. If the map is fused to much, small details at the border of focus regions could be lost. In contrary smoothing with a big kernel will increase the computation time.

The following upsampling steps inherently smooths the image, through the interpolation of missing values. Hence the map is fused with a small kernel before the upsampling step. Currently a 9 x 9 pixel flat kernel is used.



(a) (b)

Figure 8.22.: Image fusion results with smoothing of the focus map ((a) no smoothing (b) 3 x 3 flat kernel

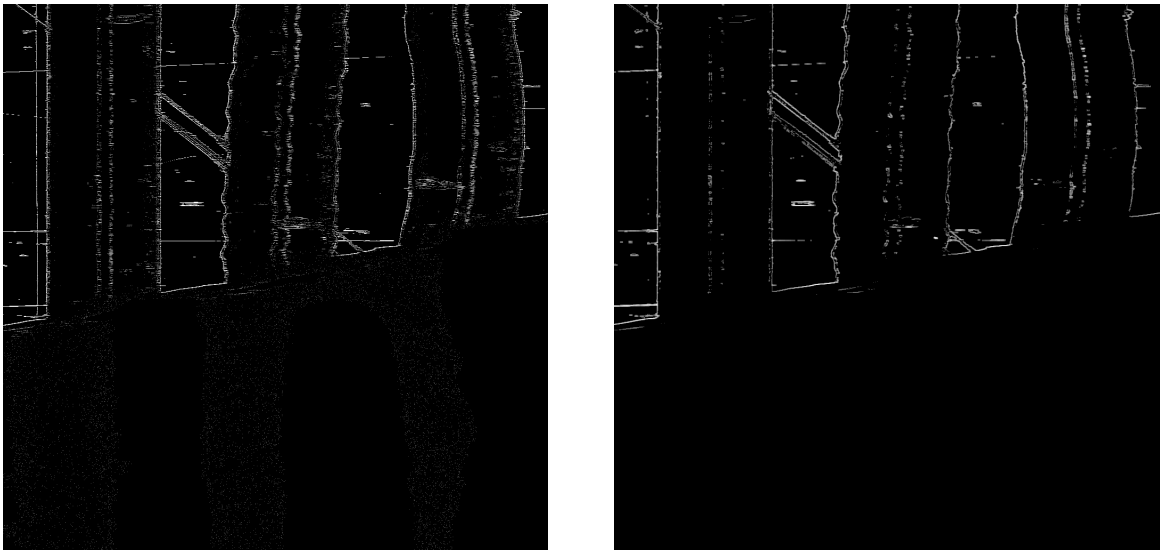
Linear combination of Pixels

TODO

Conclusion

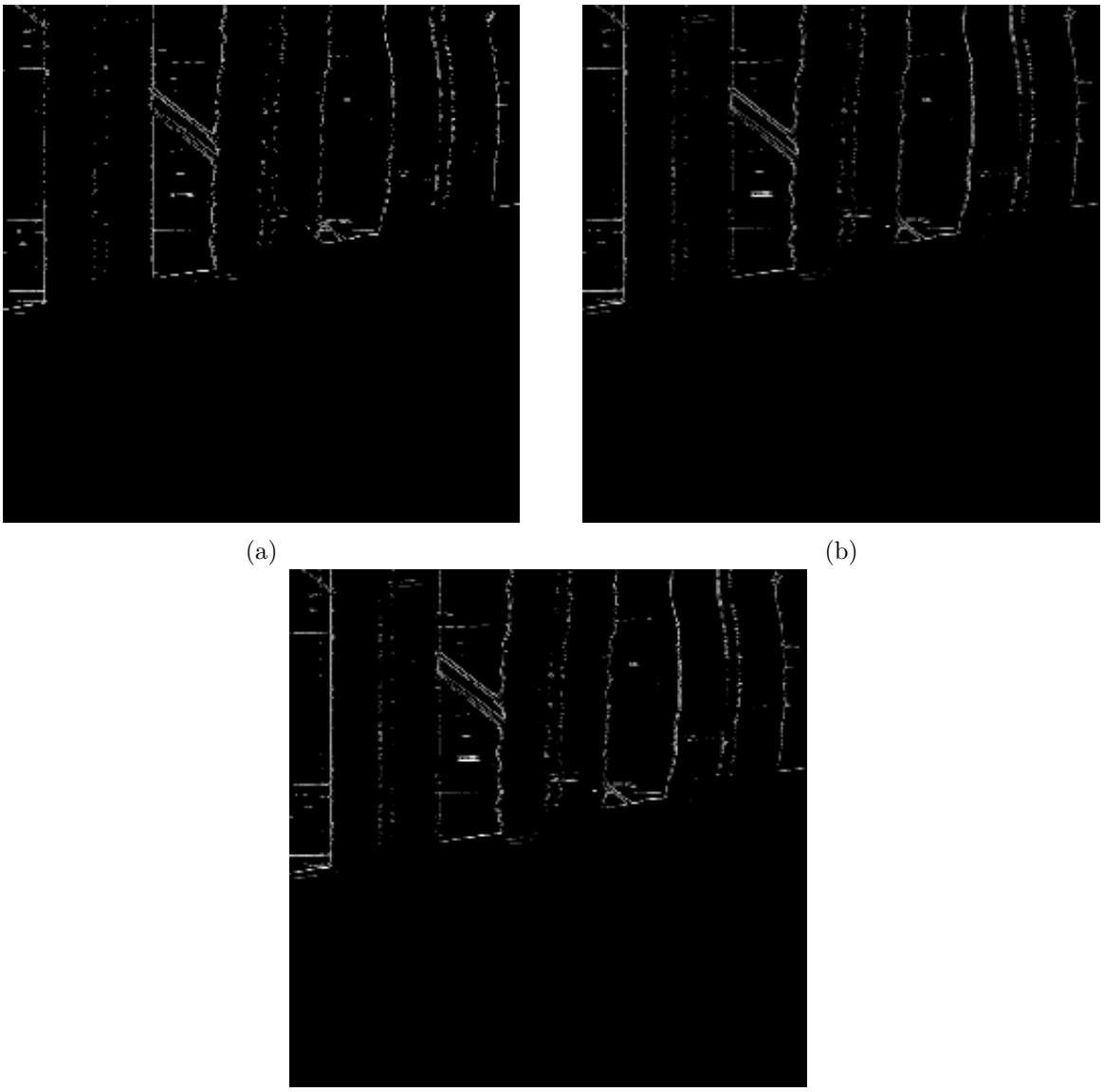
Obviously there are just to many combinations of parameters to test all possible values. Therefore the approach is to fine tune one after the other to achieve the best possible result. Figure **??** shows the output of the fusion algorithm in comparison with the mock-up created in gimp. No artifacts from the fusion process are visible in the image. The fused image provides good focus in the cracked area, although some detail is lost due to the smoothing of the focus map. However this is tolerable because the small blurred area around the crack bears very little sound information

8.3. Summary



(a) (b)

Figure 8.15.: Focus measures after thresholding ((a) Sobel, (b) morphological gradient



(a) (b)

(c)

Figure 8.16.: Comparison of interpolation methods for downsampling ((a) nearest neighbor, (b) bilinear, (c) bicubic

9. Implementation of the Fusion Algorithm

9.1. Adaptation for Large Images

Processing Flow

For the evaluation the algorithm is implemented so, that the whole focus stack is loaded at the beginning of the process. As the test images are relatively small, this is no problem. When it comes to processing the full sized scans, this is not possible because the RAM of the PCs used to run RENE is just to small. Therefore the processing flow has to be adapted.

The computer used to run RENE at the moment usually run Windows 7 and have either 8 GB or 16 GB RAM. Usually the operating system needs about 2 GB in normal operation. That leaves 6 GB for the processing of the images. The maximum size of the focus stack is assumed to be about 40 images, which is about the size that was necessary to scan the John Wolohan disc. If a 33 rpm disc is scanned one image has a file size of about 800 MB. That means up to 32 GB of data for a single rotation, obviously it is not possible to keep all these images in the RAM during the whole fusion process. That means the algorithm has to be implemented in such a way, that only several images are in the RAM at the time. The other way would be to run to process the images in parts, but that would also mean that methods to ensure smooth transition between these subimages would have to be found. Therefore it seems easier to implement the algorithm so, that the images from the stack are processed sequentially.

For the first processing step, the construction of the focus regions, this adaption is fairly easy. Instead off loading all images at the beginning, the images will be loaded only when they are needed for the construction of the fusion map. After the construction of the focus regions the focus measure values are compared against the current maximum values. Only the new values higher than the current maxima are kept and the corresponding pixels in the fusion map are updated to the index of the processed image. This implementation saves memory, because the original image and the focus regions can be released after each pass. Only the fusion map remains in the RAM during the fusion process.

After the construction of the fusion map, the actual merging of the images takes place. In the first version of the algorithm the merging consisted mainly of loop over the whole image in which each pixel is copied from the focus stack into the fused image according to the fusion map. As the whole stack can not be kept in RAM, this approach is not feasible anymore. Instead one image is loaded at the time. Then a mask is constructed from the fusion map, which shows all pixels that need to be copied. The mask is the applied to the image and the result can be added to the fused image.

Figure 9.1 shows the new process flow. The downside of this architecture is that every image has to be loaded twice from the disc, which is currently the main bottleneck.

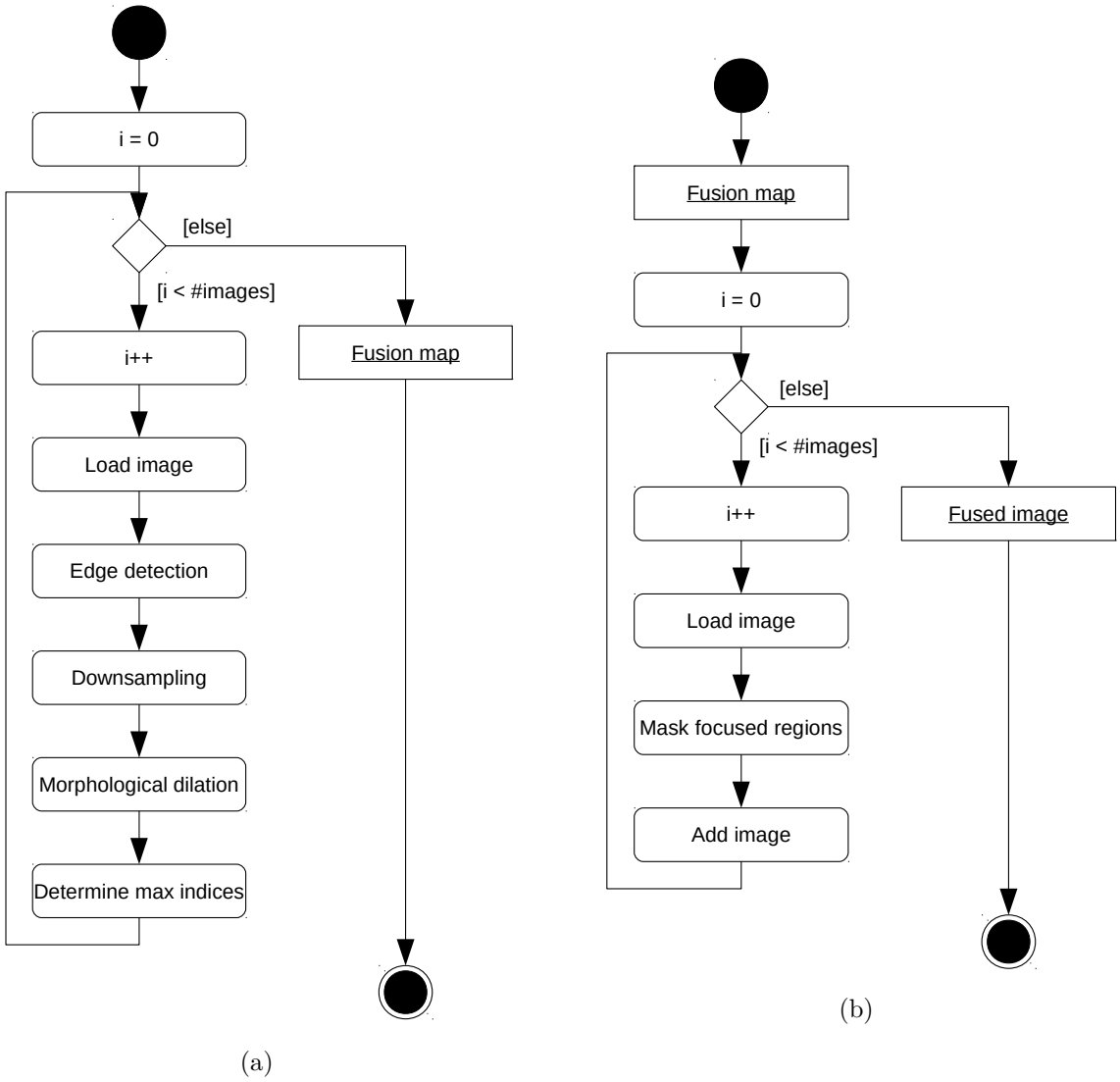


Figure 9.1.: Flow chart for the creation of the synthetic depth of field image from the fusion map

Memory Management

C# provides a sophisticated garbage collection mechanism, that means all objects that are no longer used are automatically collected and the occupied memory is released [58]. However, the time which this collection happens is not predictable. This is a problem in the fusion process, because it is not guaranteed that an image is actually released from the memory when it is no longer needed. One image can be between 300 MB to 800 MB, which means that the system RAM of 8 GB or 16 GB can be filled up relatively fast if the garbage collector can not keep up with reclaiming memory. With the first implementation of the fusion algorithm such an overflow of the RAM is observed several times. In this cases the RAM is not actually fully occupied, but the garbage collector has no free memory at hand. That means the garbage collector needs to run every time before a new object can be allocated, which slows down the whole process massively. To solve this problem a deeper look into the memory management mechanisms of C# is necessary.

One important aspect of the garbage collector is that objects that need finalization, i.e. all object that have a destructor, are not reclaimed in one step. If the GC encounters such an object it adds it to a finalization queue, which is then worked of by a dedicated thread. The finalized object will then be collected in the next run of the garbage collector.

This process can be accelerated by using the texttttIDisposable interface, which is implemented by the EMGU CV image objects [59, 60]. Calling the `Dispose` method of an object invokes it's destructor. Consequently it can be reclaimed by the GC directly without going through the finalization queue.

Calling the `Dispose` function of every image, when it is no longer need decreases the memory usage dramatically, as can be seen from the memory utilization graph **??**. This speeds up the garbage collection, but the release of memory is still nondeterministic.

TODO: add memory utilization graphs

Parallelization

In current processing flow as it can be seen in figure 9.1, only one image is processed at a time. The processing of a stack of 33 images takes about 10 min. This is a rather long processing time compared to the actual tracking and audio extraction which takes only about 1 min. Therefore the possibilities of speeding up the fusion process through parallelization are investigated.

In the first step of the fusion process the fusion map is constructed from each image. This step can easily be parallelized, by using the `ParallelFor` feature of C# [61]. This feature automatically creates threads to execute the for loop in parallel (see listing **??**). To keep the memory usage on a reasonable level, the number of parallel threads is restricted to four by setting the `MaximumConcurrencyLevel` option , which also corresponds to the number of physical CPU cores. The arrays containing the maximum values and the corresponding indices are shared between all threads. Therefore a lock needs to be added to make the update of these arrays thread safe.

```
// Construct focus region for all images
int dsFactor = 4;
Image<Gray, byte> maxVals =
    new Image<Gray, byte>(dsCols, dsRows, new Gray(0.0));
Image<Gray, byte> maxIndices =
    new Image<Gray, byte>(dsCols, dsRows, new Gray(0.0));

Parallel.For(0,
    nImages,
    new ParallelOptions { MaxDegreeOfParallelism = 4 },
    i =>
{

Image<Gray, byte> focusRegion =
    CreateFocusRegion(imageNames[i], dsFactor);
Image<Gray, byte> mask = new Image<Gray, byte>(dsCols, dsRows);
CvInvoke.cvCmp(focusRegion.Ptr, maxVals.Ptr,
                mask.Ptr, CMP_TYPE.CV_CMP_GT);

    // Update maximum values
    lock (maxIndices)
    {
        CvInvoke.cvCopy(focusRegion.Ptr, maxVals.Ptr, mask.Ptr);
        CvInvoke.cvSet(maxIndices.Ptr, new MCvScalar((double)i),
                        mask.Ptr);
    }

mask.Dispose();
focusRegion.Dispose();

});
```

Listing 9.1: Parellization of the focus region creation

The second step of the fusion process is to load each image from the stack, mask the in focus pixels and add them to fused image. This can be parallized in the same manner using a `ParallelFor` loop. In this step very little processing of the individual images is done, therefore it is questionable if the parallelization brings an actual speed up. To find out which version is preferable, both implementations are profiled.

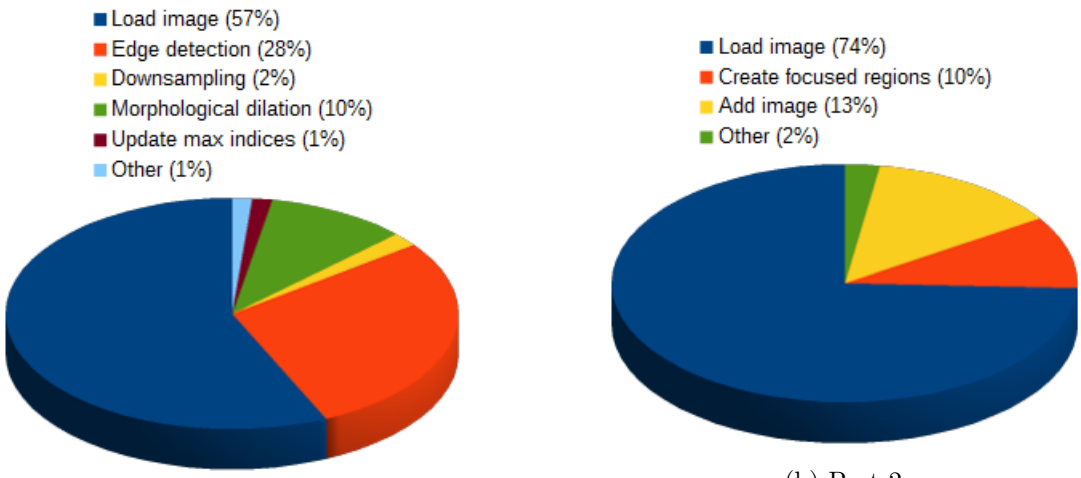
The runtime for the whole algorithm decreases only very slightly from 10 min 33 s to 10 min 29 s through the parallelization. If one looks at each steps separately the picture is different. The first steps is shortened through the parallelization from 5 min 54 s to 4 min 0 s. That means that the processing time is reduced by 31%. In contrary the processing time of the second part is increased from 4 min 39 s to 6 min 30 s. The reason for this becomes clear with a closer look at the profiling data from the single threaded variant (see figure 9.2). The profiling was done on a PC with a Intel Xeon X5472 CPU with 3.0 GHz clock speed and 16 GB RAM running Windows 7 Professional SP1, 64 Bit. The images are stored on a Western Digital WDC40EFRX 4 TB hard drive which is connected to a SATA revision 1.0 interface, which both have a maximal transfer rate of 150 MB/s.

For the first processing step the loading of the image needs 57% of the processing time, for the second part 74%. The time needed to load the image is mainly given by the transfer rate of the hard disc, which can only be accessed by one thread at at time. If several threads try to access the hard disc at the same time all except one are blocked and they have to wait for their turn

to load the image. Therefore the parallelization only makes sense for the first part of the fusion process, which needs a higher portion of the run time for actual processing of the image.

Processing step	Serial execution	Parallel execution
1	5 min 54 s	4 min 0 s
2	4 min 39 s	6 min 30 s
Total	10 min 33 s	10 min 29 s

Table 9.1.: Execution times of the serial and parallel variants of the image fusion algorithm



(a) Part 1

(b) Part 2

Figure 9.2.: Profiling results for the serial variant of the image fusion algorithm

9.2. Integration in RENE

The image fusion process was implemented in a single class, named `ImageFuser`, which is added to the RENE code. Depending on which tracking mode is used in RENE, the data flow can vary substantially. The majority of tracking methods is implemented simply in functions which take the image array as parameters. Therefore the image fusion step can just be added between the loading of the image and the tracking. To allow the user to chose between focus stacking and normal operation a new check box was added to the GUI (see figure 9.3).

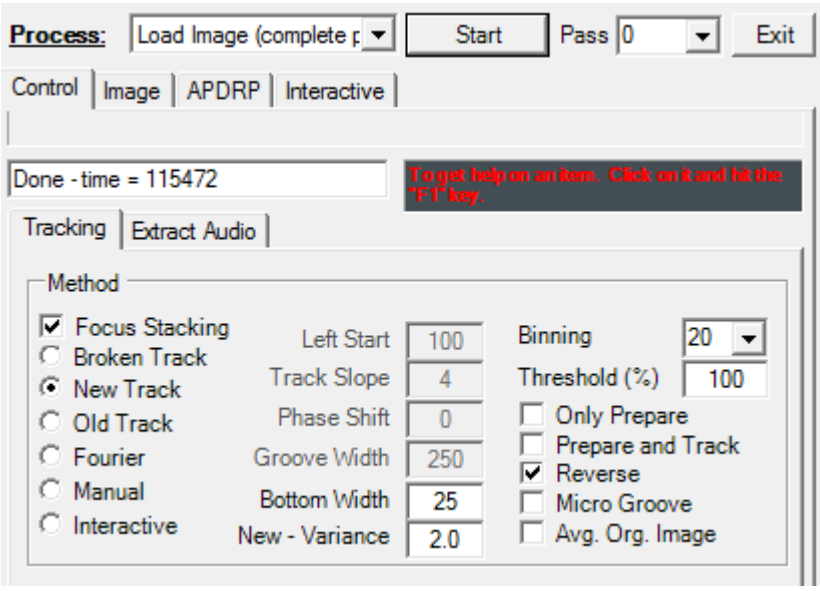


Figure 9.3.: Screenshot of the new RENE user interface

The *Broken Track* method is the exception, it is implemented as class with the image loading as member function. As there is no way to pass an image to this class, a new track handler class has to be developed which allows the usage of the *Broken Track* method together with the focus stacking. The new class inherits all functionalities from the `BrokenTrackHandler`, except the image loading function. The new image loading function creates an `ImageFusion` object and runs the fusion process. The architecture of the new `FusionTrackHandler` class can be seen in figure 9.4.

9.3. Summary

Several steps are necessary to adapt the fusion algorithm for processing full sized images of 300 MB to 800 MB. The processing flow is adapted to reduce the RAM usage. In addition the memory management needed to be improved. After these refactoring steps the algorithm is integrated in the RENE software.

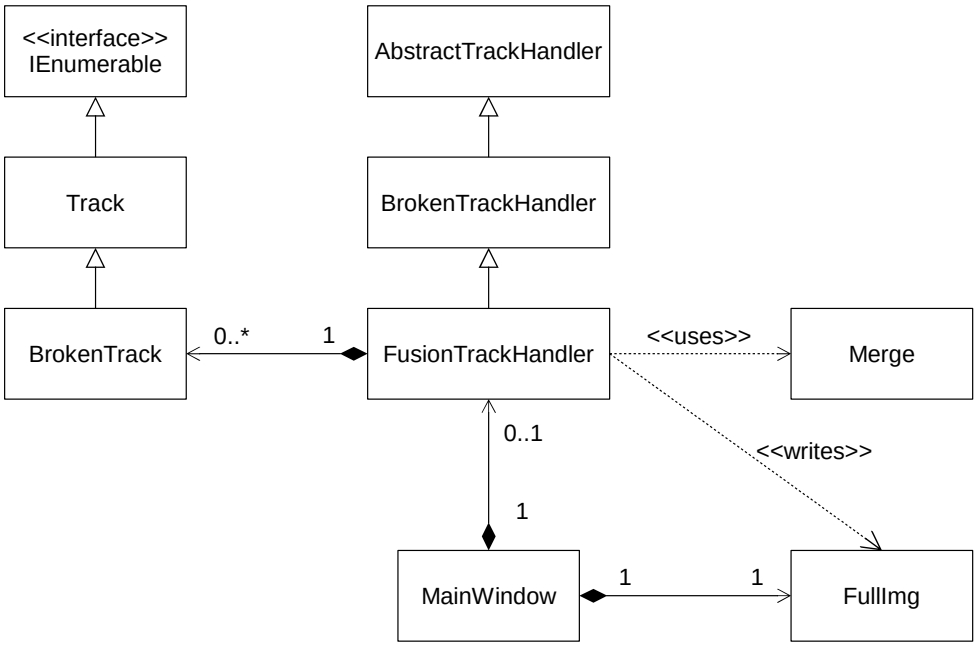


Figure 9.4.: Class diagram of the tracking algorithm

Part IV.

Tests and Conclusion

10. Tests and Verification

10.1. Disc in Good Condition

This test was made to find out if the fusion process affects the audio quality or if it introduces artifacts in the extracted signals. A disc in good condition is scanned with the tracking and the focus stacking method and the extracted audio is then compared. As the disc does not contain any cracks, the standard tracking methods can be used. This test should provide an indication how well the fusion algorithm handles the warpage of a disc. As test object the MC-101-3 disc, which was already used for the focus experiment, is chosen as test object.

10.1.1. Disc Warpage

The first step is to determine the radius of the disc with the highest warpage. Therefore the warpage of the disc is determined at 3 positions, the outermost, the middle and the innermost part of the groove. Table 10.1 shows the measurement results. The warpage is clearly highest at the outermost radius of the groove, thus this part of the disc is used for the test (figure 10.1 shows the height profile for this part). A scan containing two revolutions from 119 mm to 117 mm is used for this test. The first revolution consist to one half of the silent lead in groove, the second revolution is fully occupied with music.

Radius	**Warpage**
119 mm	124 µm
89 mm	103 µm
59 mm	37 µm

Table 10.1.: Warpage of the MC-101-3 lacquer disc at different radii

10.1.2. Visual Comparison

To check if there are artifacts in the fused image it is inspected visually and compared against the scan done with the tracking method.

The first overall impression is that the fused image provides more detail. As it can be seen in figure 10.2 the fused image provides a good focus for the groove top and bottom, were as on the image made with the tracking method the groove top is slightly blurred.

At some points in the fused image visible artifacts occur at the groove top or bottom. The artifacts at the groove bottom are visible as sudden changes in the width of the groove bottom (see figure 10.3). The artifacts at the groove bottom appear as bits of the groove top missing (see figure 10.4). Telling from the look of the artifacts it looks like they are produced through small changes in the fusion map. Thus a small area is filled with pixels from another image from the stack. These patches could may be lessened by reducing the granularity of the fusion map,

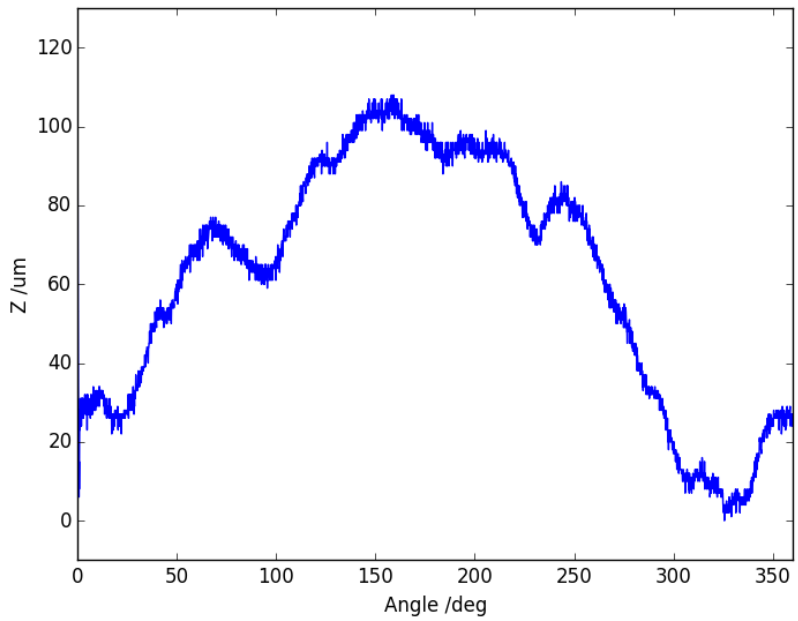


Figure 10.1.: Height profile of the MC-101-3 lacquer disc at radius 119 mm

(a) Focus stacking method

(b) Tracking method

Figure 10.2.: Comparison of an image made with the focus stacking method against one made with the tracking method

through a higher downsampling factor, bigger structuring elements in the dilation or smoothing with a bigger kernel. But this would probably also have a negative impact on areas around cracks. The artifacts are relatively small, but they could still introduce errors in the audio.

10.1.3. Audio Comparison

The audio extracted from the second scanning revolution is compared, the section is about 8 seconds long. Although both scans are made a the same radius, there can still be a slight offset between the two tracks. Therefore a crosscorelation of the first 100'000 samples is made to determine the offset between the signals and align the tracks.

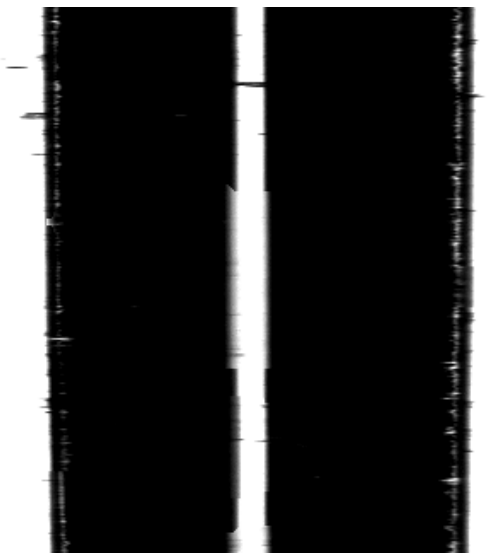
Figure 10.5 shows the waveform for both tracks. If the signals are compared directly, the differences seems rather small. However, a relatively big error remains if the difference between both signals is calculated (see figure 10.6). The fact that the error seems not to change in parallel to the signal amplitude indicates that it could be mainly noise. In the frequency domain the difference is better visible. The spectrum from the focus stacking signal is about 1 dB lower in the low frequency part up to 100 Hz but about the same amount higher in the high frequency part above 4 kHz 10.7. As the difference is lower than the just-noticable difference of 1 dB for the whole frequency range, it can be assumed that the difference is not audible even for a trained ear. The spectra are calculated by averaging FFTs with a block size of 8192 samples with the same tools as explained in section 5.2.

Periodic Artifacts

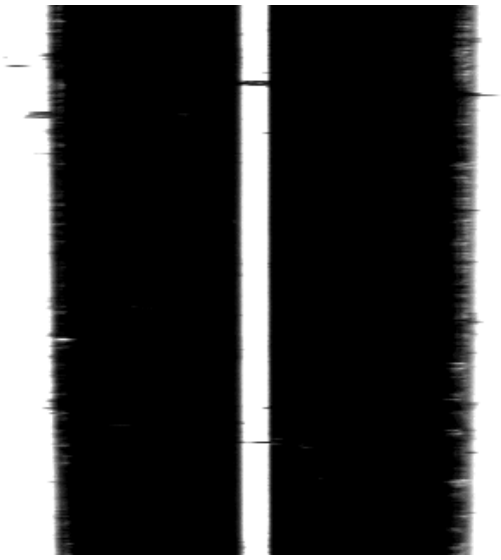
To see if the error signal contains any periodic artifacts, it's spectrum is calculated. The idea behind is that if the fusion process creates artifacts in the image, they might show at certain frequencies. For example the downsampling rate of 4 could create a pattern at about quarter of the sampling frequency, i.e. 55 kHz. Figure 10.8 shows the spectrum of the error signal. Interestingly increases the spectrum with the frequency, with a maximum at about 4 kHz. The reason for that is not clear.

Another approach to search for artifacts introduced through the fusion process is too look at the audio extracted from the lead in groove of the disc. As already mentioned in section 5.2, does the lead in groove contain no audio. The groove at this section appears fully straight in the image, which means that the extracted audio contains only noise. If the fusion process creates periodic artifacts it is likely that they show up in this part, because they would not be covered by the audio signal. Obviously can two section containing only noise not be aligned with a crosscorelation. Therefore the first 100'000 samples after the start of the music part are used to align the signals.

Figures 10.9 and 10.10 show the signals waveforms and the difference signals for both methods. In this comparison the focus stacking signal appears to contains more noise that the one extracted from the tracking scan. This is confirmed by the spectra of booth signals. The spectrum of the focus stacking signal is higher over the whole range and it contains especially more noise in the middle frequencies (see figure 10.11). However there is no significant peak in the spectrum which would indicate a periodic artifact. The spectra of the difference signal again increase over the frequency range, similar to the previous test (see figure 10.12).

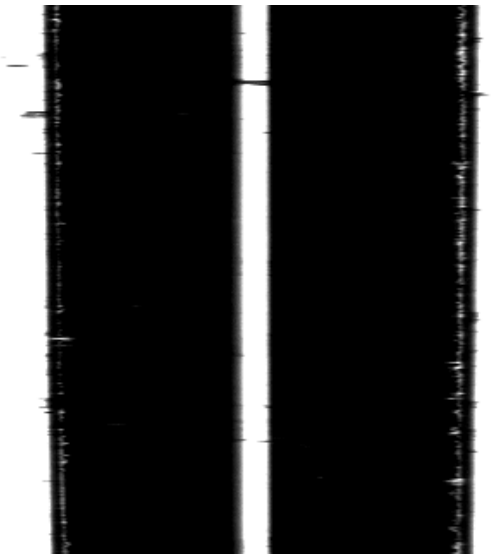


(a) Focus stacked image



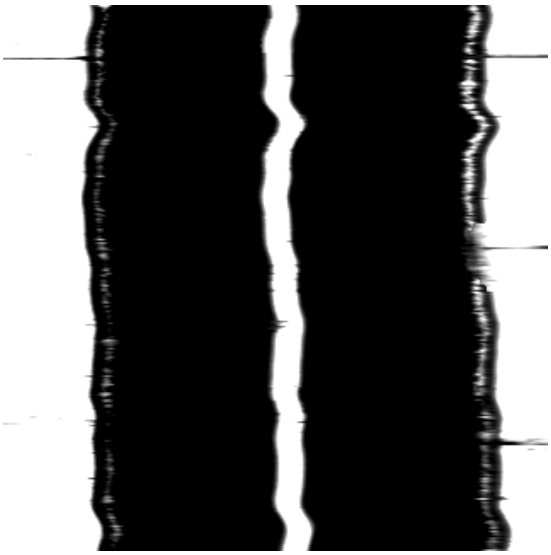
(b) Tracking image

(c) Image 2/6 from the focus stack



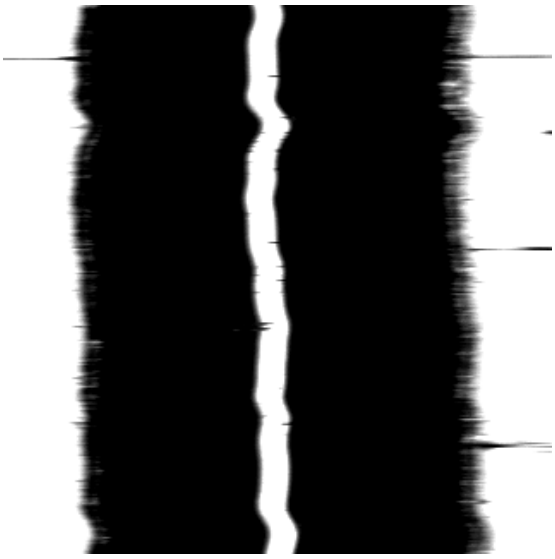
(d) Image 4/6 from the focus stack

Figure 10.3.: Artifacts at the groove bottom

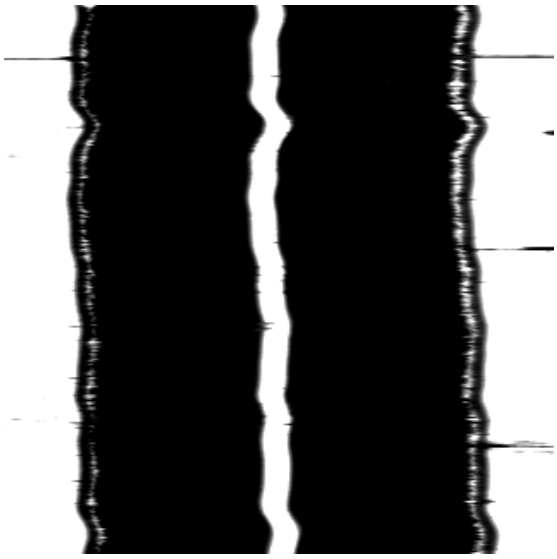


(a) Focus stacked image

(b) Tracking image



(c) Image 3/6 from the focus stack



(d) Image 4/6 from the focus stack

Figure 10.4.: Artifacts at the groove top

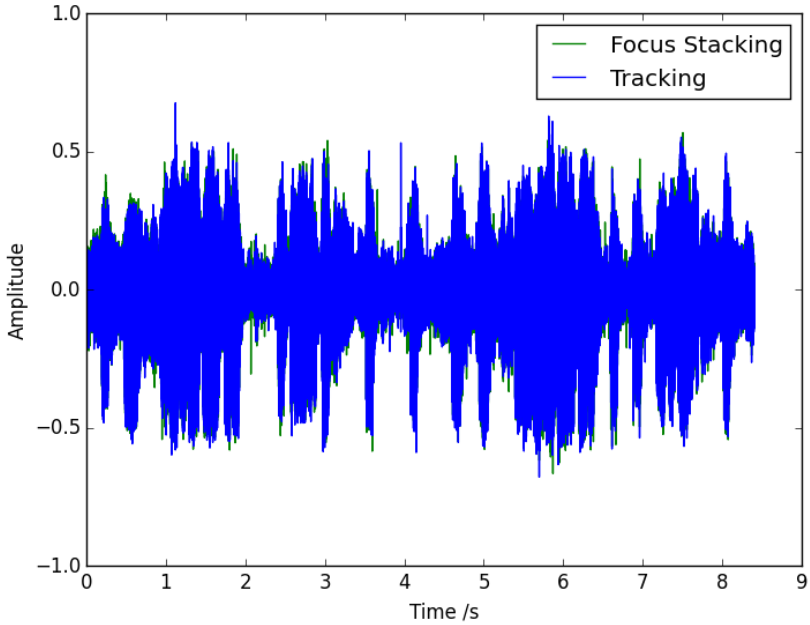


Figure 10.5.: Comparison of the signal extracted with the focus stacking and tracking method

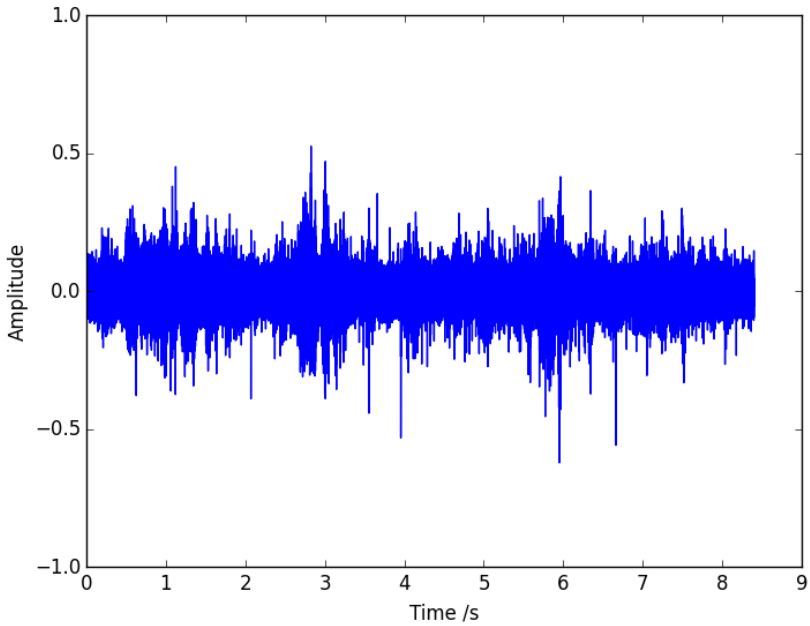


Figure 10.6.: Difference between the audio extracted with the tracking and the fusion method

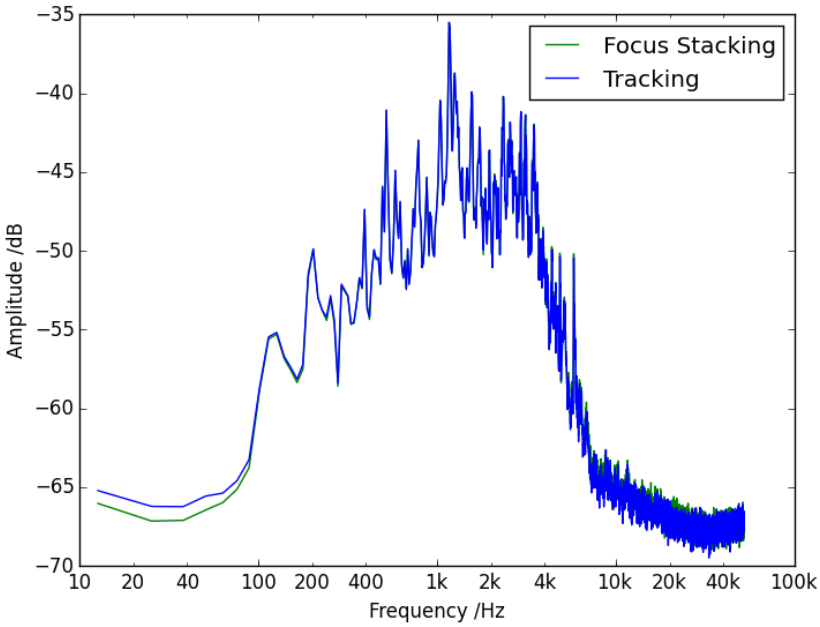


Figure 10.7.: Comparison of the spectra of the signal extracted with the focus stacking and tracking method

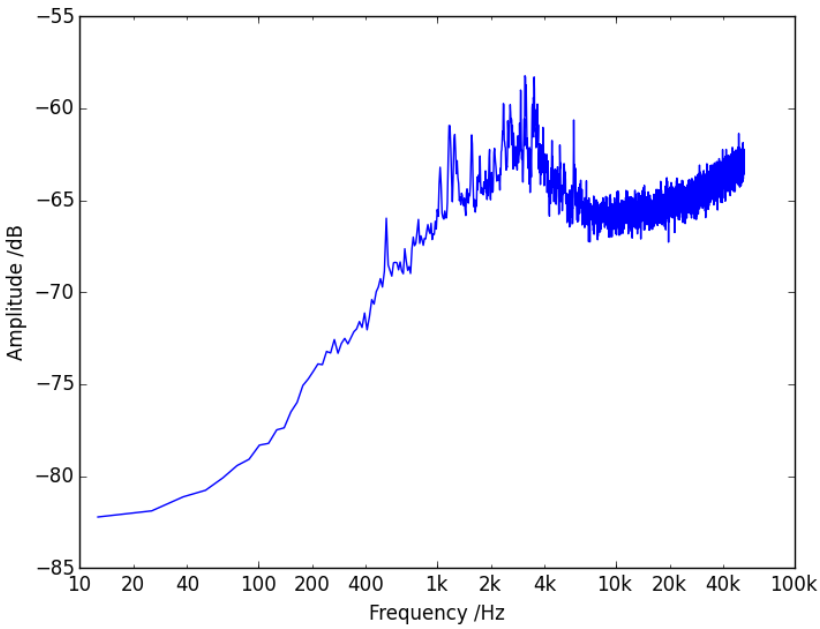


Figure 10.8.: Spectrum of the difference signal from the audio extracted with the tracking and the fusion method

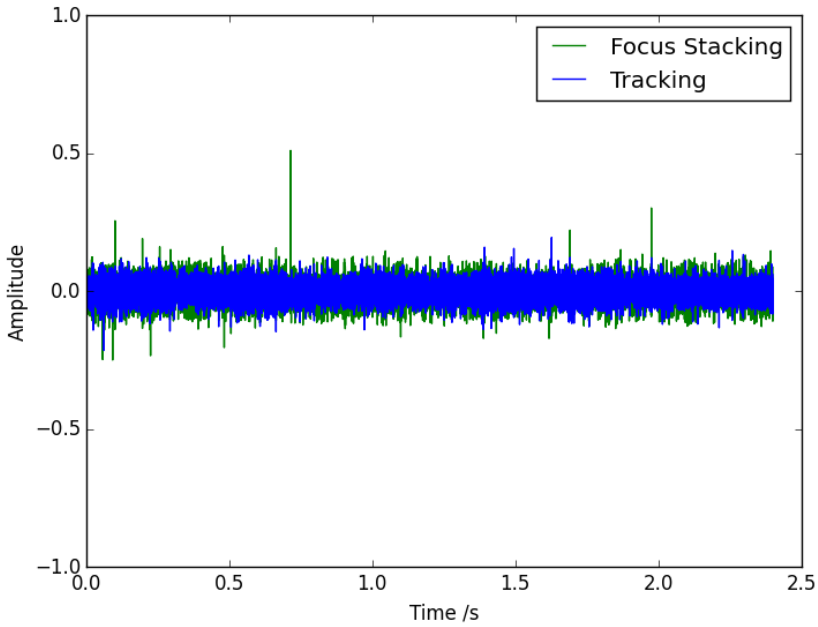


Figure 10.9.: Comparison of the signal extracted from the lead in groove with the focus stacking and tracking method

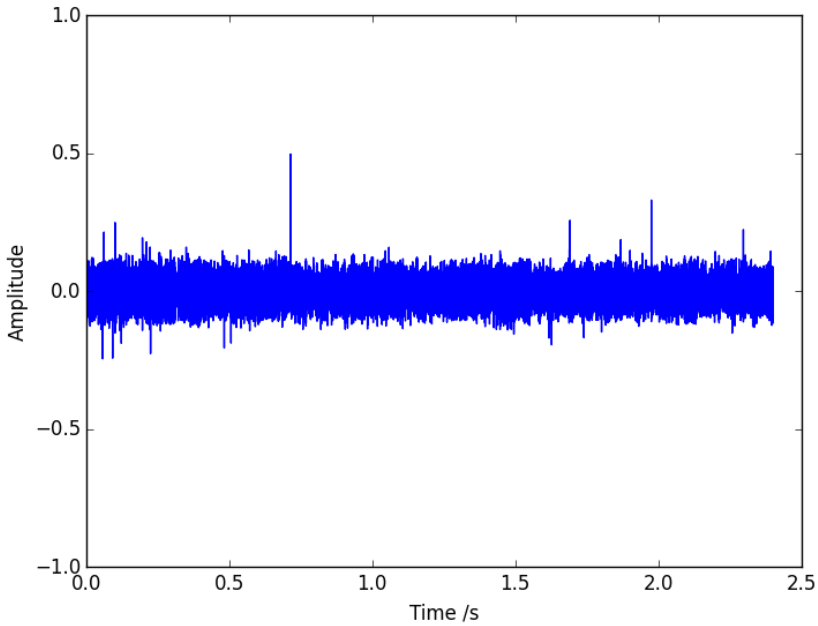


Figure 10.10.: Difference between the audio extracted from the lead in groove with the tracking and the fusion method

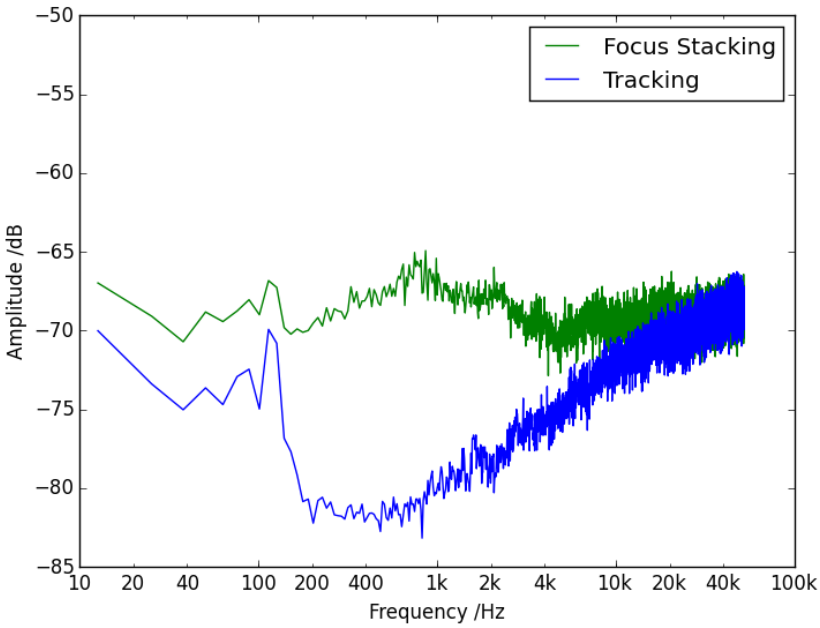


Figure 10.11.: Comparison of the spectra of the signal extracted from the lead in groove with the focus stacking and tracking method

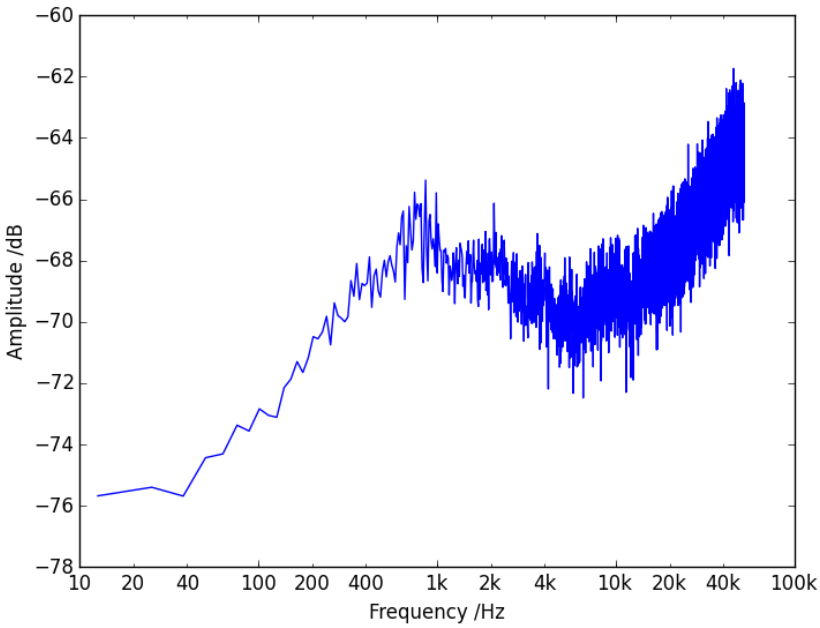


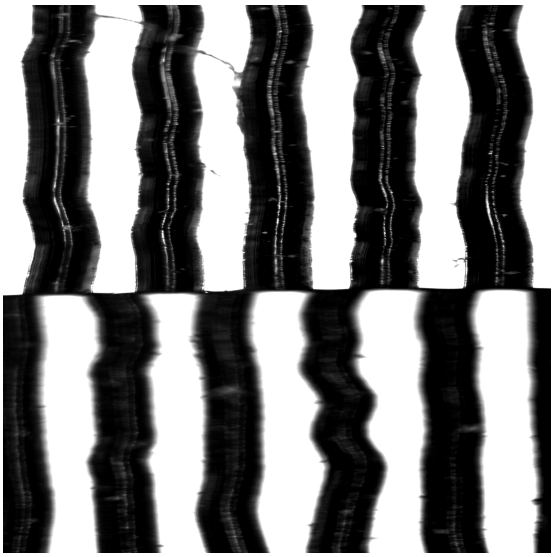
Figure 10.12.: Spectrum of the difference signal from the audio extracted from the lead in groove with the tracking and the fusion method

10.1.4. Delaminating Disc

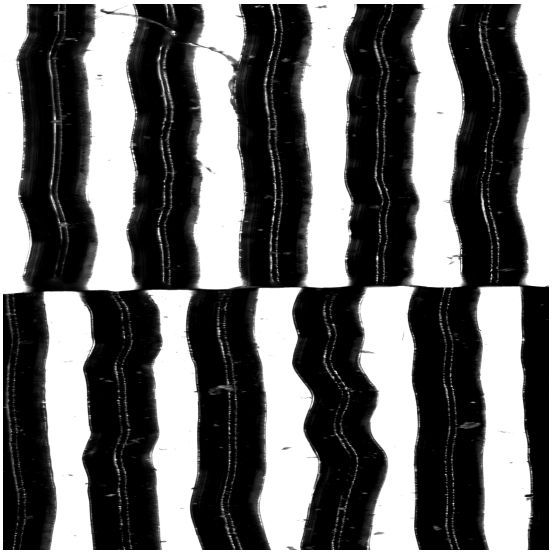
This test is made to see how the new method compares to the tracking method when it comes to cracks in delaminating lacquer disc. As test object the John Wolohan disc is used, which was already used for the test image. The disc shows the delamination that is typical on lacquer discs. For the test the innermost section is scannend, because it shows the most damage from delamination.

10.1.5. Visual Comparison

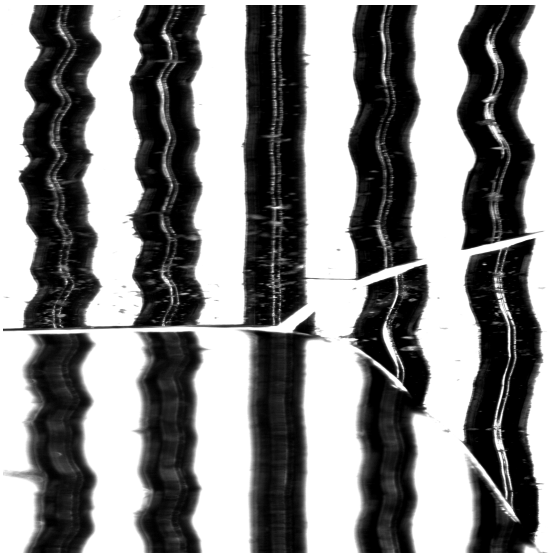
As first step the scans are compared visually.



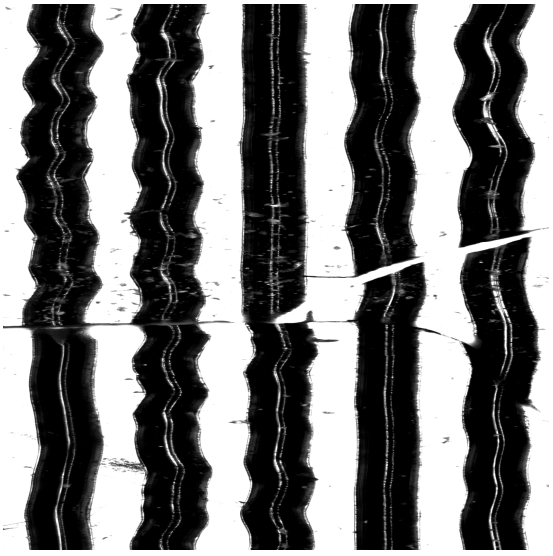
(a) Tracking



(b) Focus stacking



(c) Tracking



(d) Focus stacking

Figure 10.13.: Comparison of areas around cracks from the tracking and the focus stacking method

10.1.6. Audio Extraction

In the first attempt to extract audio, the *BrokenTrack* method is used, to see if the automatic relinking of the track can be used on this record. However tracking of the grooves with this method is not possible. The method has several parameters that can be adjusted, but even after several runs with different setting not tracking is possible. The problem is that the method applies a binary threshold to the image, before the tracking. The John Wolohan disc has a very fine

Two sections of the disc are scanned. The outermost part which is more ore less intact and contains only two cracks and the innermost part of the disc which shows severe damages through delimination. For comparison the scans are made with the tracking and the focus stacking method.

The scan of the outermost part provides a reference point for how the audio of the delaminated part would sound like, if the focus over the cracks can be maintained and the tracks are relinked well. Figure 10.14 shows the spectra for both acquisition methods. The spectra look very similar, except that the focus stacking method produces more high freqcuency content. Judging from the hearing impresion this seems to be mainly noise.

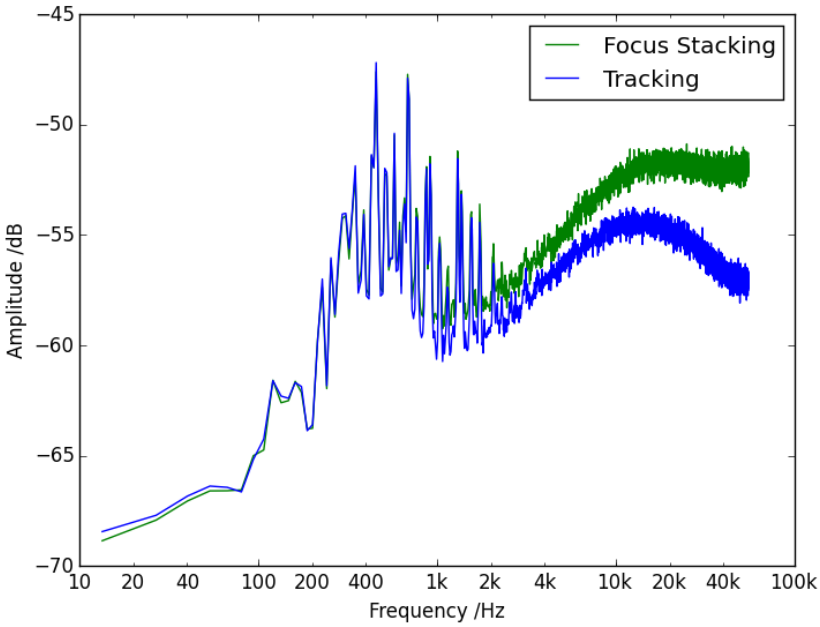


Figure 10.14.: Comparison of the spectra from the audio extracted from the outside part of the John Wolohan disc

In the second step the innermost part of the disc is scanned. The intention of this test is to determine if the loss of focus at the cracks is visible in the spectrum. The loss of focus in the areas around cracks means that small details get lost, which could create a attenuation effect similar to a low pass filter. The blurred regions are usually rather small, that means that this effect is probably only detectable on sections with a high number of cracks. The innermost part of the John Wolohan disc is used for this test, because it contains a high number of cracks. They can be clearly seen as sudden changes in the height profile (see figure 10.15). If there is no such effect, the spectra should similar to the one from the outside of the disc.

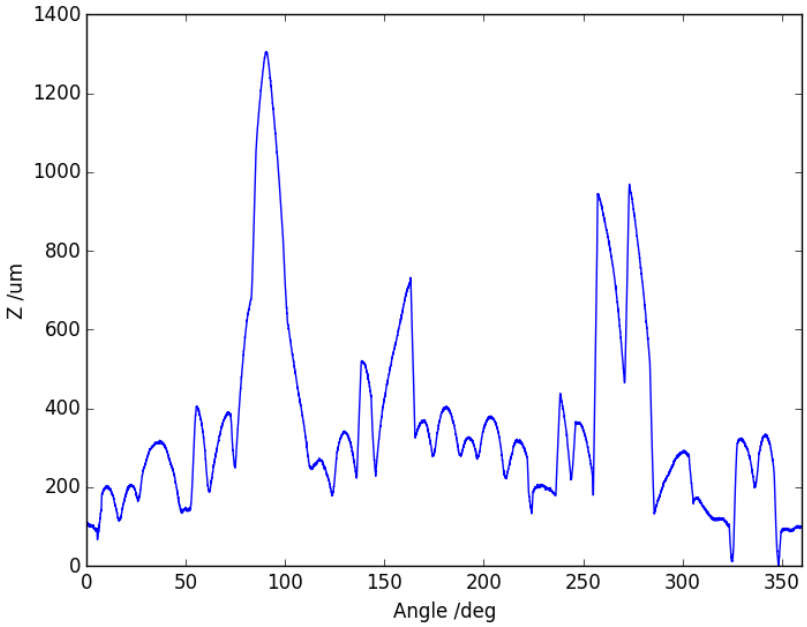


Figure 10.15.: Height profile of the innermost section of the John Wolohan disc

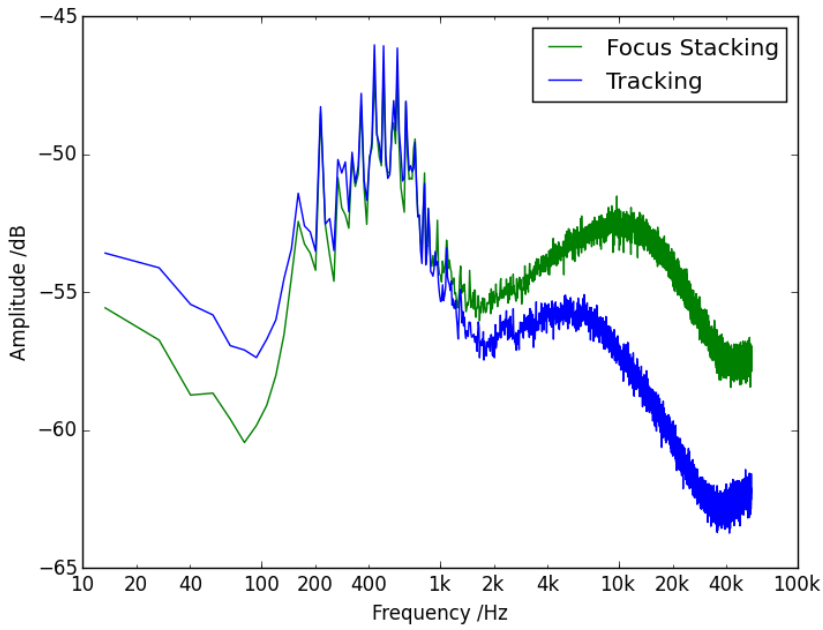


Figure 10.16.: Comparison of the spectra from the audio extracted from the inside part of the John Wolohan disc

The spectrum of the audio extracted from the tracking scan shows an attenuation above 2 kHz. This indicates that

10.2. Summary

11. Conclusion and Further Work

11.1. Conclusion

11.1.1. Image acquisition

Acquisition system is successfully upgraded

Depth of field of the current optical system determined. The result coincides with the theoretical estimation. The reason why the RMS is not symmetrical around the best focus point has to be found? Maybe the focus point is not optimal (the focusing method should be further investigated). Or maybe the focus is influenced by the fact that the groove bottom is about 70 um apart from the top has an influence. At the moment it is unclear what is actually focused groove top or bottom.

The acquisition system works but the size of the focus stack is high, that leads to long scanning times and a lot of data. It would be worth while investigating if the depth of field of the system can be enlarged through changes in the optical system. Adapting the camera trajectory to reduce the number of scans could also be a solution. However the the approximation must be made in a way, so that no information about the cracked regions is lost.

11.1.2. Image fusion

The current algorithm delivers good results. Some artifacts are still present in the fused image. Therefore there is still a lot of room for optimization. The runtime could be reduced through parallelization or GPGPU utilization. A good number of approaches have been evaulated, but the evaulation covered only spatial domain methods. Because the current algorithm has already some complexity wavelet methods might be worth while investigating

The image fusion algorithm is integrated in the RENE software for audio extraction. However currently the algorithms use completly different interfaces. For example is at not possible to use the relinking Methods from BrokenTrack but the Fourier method for tracking. It would be desirable to have standartized interfaces between the processing steps, so that the tracking methods could be converted to interchangable methods. That would allow to swap out processing steps, so that the best suited method for the current disc can be found more quickly.

The algorithm delivers good image quality around cracked regions. However the relinking method is missing at the moment. Fischers method fails for a lot of discs, because it uses a global threshold. It should be investigated if local thresholding method could increase the robustness.

11.2. Further Work

The investigation of the noise RMS vs. Z-offset showed that a higher uncertainty of the groove edge position creates a lof of noise in the audio extraction. Currently the determined groove

edge position is directly used for the audio extraction. With prediction filtering technicques like Kalman filtering the audio extraction could may be improved.

* Use Kalman filters to improve the audio extraction on blurred edges
* Try to apply compressed sampling to the focus problem
* Use lenses with ultrasonic motors to speed up the acquisition
* From the focus information depth information can be derived (depth from focus) which could be used to improve the audio extraction

Glossary

FFT	Fast Fourier Transform. An algorithm to compute the discrete Fourier transform of a signal, used for spectral anlysis [**?**].
InES	Institute of Embedded Systems at the ZHAW.
IRENE	Image, Reconstruct, Erase Noise, Etc. The name of the sound restoration project at the LBNL using an optical scanner.
IRENE 2D	Name of the scanning system for disc recordings using a 2D camera in the IRENE project.
IRENE 3D	Name of the scanning system for cylinder recordings using a 3D probe in the IRENE project.
JND	Just-Noticable Difference. The minimal increase in amplitude of an audio signal that can be perceived by humans [36].
LBNL	Lawrence Berkeley National Laboratory
PRISM	Name of the software used for audio extraction from 3D scans in the IRENE project.
RENE	Name of the software used for audio extraction from 2D scans in the IRENE project.
ZHAW	Zurich University of Applied Sciences

Bibliography

[1] Wikipedia. Édouard-Léon Scott de Martinville. http://en.wikipedia.org/wiki/%C3%89douard-L%C3%A9on_Scott_de_Martinville [03/08/2015].

[2] S. C. Cavaglieri, O. Johnsen and F. Bapst. Optical Retrival and Storage of Analog Sound Recordings. In *Audio Engineering Society 20th International Conference*, page 1948, October 2001.

[3] Northeast Document Conservation Center. IRENE Seeing Sound Blog. A Hefty Challenge for IRENE: Working with Delaminating Lacquer Discs, August 2014. https://www.nedcc.org/audio-preservation/irene-blog/[03/14/2015].

[4] Stephanie A. Hall. Save Our Sounds: America's Recorded Sound Heritage Project, January 2015. http://www.loc.gov/folklife/sos/preserve1.html [03/04/2015].

[5] The Engineering ToolBox. Coefficients of Linear Thermal Expansion. http://www.engineeringtoolbox.com/linear-expansion-coefficients-d_95.html [06/10/2015].

[6] Mitutoyo. Microscope Units and Objectives (UV, NUV, Visible & NIR Region). Catalog No. E4191-378. http://www.mitutoyo.com/wp-content/uploads/2012/11/E4191-378_010611.pdf [05/24/2015].

[7] S. Fischer. Restoring Cracked Early Recordings. Master's thesis, Zurich University of Applied Sciences, September 2013.

[8] V. Fadeyev and C. Haber. Reconstruction of Mechanically Recorded Sound by Image Processing. *Journal of the Audio Engineering Society*, 51(12):1172 – 1185, December 2003.

[9] J. Singy. Cracked Records with 3D/IRENE. Master's thesis, University of Applied Sciences Western Switzerland, February 2013.

[10] Atlassian. Comparing Workflows. https://www.atlassian.com/git/tutorials/comparing-workflows/ [07/09/2015].

[11] Mickey Hart. Preserving Our Musical Heritage: A Musician's Outreach to Audio Engineers. *Journal of the Audio Engineering Society*, 49(7/8):667–670, July 2001.

[12] Wikipedia. Phonautograph. http://en.wikipedia.org/wiki/Phonautograph [05/23/2015].

[13] First Sounds. Édouard-Léon Scott de Martinville's Phonautograms. http://www.firstsounds.org/sounds/scott.php [05/23/2015].

[14] First Sounds. Fixation et Transcription du Chant (1860) - Édouard-Léon Scott de Martinville, note = http://www.firstsounds.org/publications/facsimiles/firstsounds_facsimile_05.pdf [03/08/2015].

[15] Wikipedia. Phonograph. http://en.wikipedia.org/wiki/Phonograph [03/08/2015].

[16] Library of Congress. Thomas Edison. http://www.loc.gov/pictures/item/brh2003000454/PP/ [05/23/2015].

[17] Wikipedia. Gramophone Record. http://en.wikipedia.org/wiki/Gramophone_record [03/16/2015].

[18] Wikipedia. Transcription disc. http://en.wikipedia.org/wiki/Transcription_disc [06/04/2015].

[19] Wikipedia. Acetate Disc. http://en.wikipedia.org/wiki/Acetate_disc [05/24/2015].

[20] The 78 Project. Getting to Know the PRESTO #1. Nightmares in Recording, and in History. http://the78project.com/ getting-to-know-the-presto-1-nightmares-in-recording-and-in-history/ [05/24/2015].

[21] Newport. *XPS-Q8: Universal High-Performance Motion Controller/Driver. User's Manual, Software Tools and Tutorial*, March 2015. V1.4.x.

[22] Newport. XM Series: Ultra-Precision Linear Motor Stages. http://assets.newport.com/webDocuments-EN/images/XM_Data_Sheet_MC.pdf [03/30/2015].

[23] Newport. LTA Series: Precision Motorized Actuators. http://assets.newport.com/webDocuments-EN/images/LTA_Data_Sheet_MC.pdf [03/30/2015].

[24] Newport. RGV100: High-Speed Precision Rotation Stages. http://assets.newport.com/webDocuments-EN/images/RGV100_Data_Sheet.pdf [03/30/2015].

[25] Lumencor. SOLA light engines: Solid-State White Light Excitation Subsystems. http://lumencor.com/wp-content/uploads/sites/3/2014/09/1553-LUM. SOLA-54-10021REV2.pdf [03/30/2015].

[26] Keyence. LKG-157. http://www.keyence.com/products/measure/laser-1d/lk-g3000/models/lk-g157/index.jsp [03/30/2015].

[27] Basler. *Basler Racer. User's Manual for GigE Vision Cameras*, September 2013. Version 3.

[28] Wikipedia. Sampling (signal processing). http://en.wikipedia.org/wiki/Sampling_%28signal_processing%29 [05/25/2015].

[29] Teledyne Dalsa. Piranha 2 P2-42-04K40 (7 um). Line Scan Cameras. https://www. teledynedalsa.com/prot/mv/manuals/00493-14_03-32_Piranha2_User_Manual.pdf [05/29/2015].

[30] National Instruments. NI Vision Acquisition Software. Software to Acquire, Save, and Display Images. http://sine.ni.com/nips/cds/view/p/lang/en/nid/12892 [05/29/2015].

[31] D. B. Murphy and M. W. Davidson. *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley-Blackwell, Hoboken, NJ, 2nd edition, 2013.

[32] K. R. Spring and M. W. Davidson. Nikon. Depth of Field and Depth of Focus. https://www.microscopyu.com/articles/formulas/formulasfielddepth.html [07/21/2015].

[33] E. Jones, T. Oliphant, P. Peterson and others. SciPy: Open Source Scientific Tools for Python, 2001. http://www.scipy.org [03/29/2015].

[34] S. van der Walt, S. C. Colbert and G. Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13(2):22 – 30, March - April 2011.

[35] J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing In Science and Engineering*, 9(3):90 – 95, May - June 2007.

[36] Wikipedia. Just-Noticeable Difference. http://en.wikipedia.org/wiki/Just-noticeable_difference [06/04/2015].

[37] Thor Labs. Resolution Test Targets. https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=4338&pn=R1L3S10P [06/22/2015].

[38] H. Mir, P. Xu and P. van Beek. An Extensive Empirical Evaluation of Focus Measures for Digital Photography. In *Digital Photography X*, March 2014.

[39] S. Yousefi, M. Rahman, N. Kehtarnavaz, M. Gamadia. A New Auto-Focus Sharpness Function for Digital and Smart-Phone Cameras. In *2011 IEEE International Conference on Consumer Electronics (ICCE)*, pages 475 – 476, January 2011.

[40] A. Santos, C. Ortiz de Solórzano, J. J. Vaquero, J. M. Peña, N. Malpica and F. del Pozo. Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of Microscopy*, pages 264 – 272, December 1997.

[41] R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Pearson, Upper Saddle River, NJ, 3rd edition, 2010.

[42] OpenCV. Sobel Derivatives. http://docs.opencv.org/doc/tutorials/imgproc/imgtrans/sobel_derivatives/sobel_derivatives.html [06/29/2015].

[43] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, January 2000.

[44] G. Bradski and A. Kaehler. *Learning OpenCV. Computer Vision with the OpenCV Library.* O'Reilly, Sebastopol, CA, 1st edition, September 2008.

[45] R. J. Wall, S. S. Sobin, M. Karspeck, R. G. Lindal, H. M. Tremer and Y. C. Fung. Computer-Derived Image Compositing. *Journal of Applied Physiology*, 51(1):84 – 89, July 1981.

[46] A. G. Valdecasas, D. Marshal, J.M. Becerra and J. J. Terrero. On the Extended Depth of Focus Algorithms for Bright Field Microscopy. *Micron*, 32(6):559 – 569, August 2001.

[47] R. J. Pieper and A. Korpel. Image processing for extended depth of field. *Applied Optics*, 22(10):1449 – 1453, May 1983.

[48] H. A. Eltoukhy and S. Kavusi. A Computationally Efficient Algorithm for Multi-Focus Image Reconstruction. In *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications IV (SPIE 5017)*, pages 332 – 341, May 2003.

[49] H.-S. Wu, J. Barba and J. Gil. A Focusing Algorithm for High Magnification Cell Imaging. *Journal of Microscopy*, 184(2):133 – 142, November 1996.

[50] S. Li, J. T. Kwok and Y. Wang. Combination of Images with Diverse Focuses Using the Spatial Frequency. *Information Fusion*, 2(3):169 – 176, September 2001.

[51] W. Huang and Z. Jing. Evaluation of Focus measures in Multi-Focus Image Fusion. *Pattern Recognition Letters*, 28(4):493 – 500, March 2007.

[52] S. Li, B. Yang. Multifocus Image Fusion Using Region Segmentation and Spatial

Frequency. *Image and Vision Computing*, 26(7):971 – 979, July 2008.

[53] H. Li, B. S. Manjunath and S. K. Mitra. Multi-Sensor Image Fusion Using the Wavelet Transform. *Graphical Models and Image Processing*, 57(3):235 – 245, May 1995.

[54] R. Polikar. The Wavelet Tutorial. The Engineer's Ultimate Guide to Wavelet Analysis, January 2001. http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html [07/15/2015].

[55] C. Valens. A Really Friendly Guide to Wavelets, October 2010. http://polyvalens.pagesperso-orange.fr/clemens/wavelets/wavelets.html [07/15/2015].

[56] Y. Song, M. Li, Q. Li and L. Sun. A New Wavelet Based Multi-focus Image Fusion Scheme and It's Application on Optical Microscopy. In *Proceedings of the 2006 IEEE International Conference on Robotics and Biomimetics (ROBIO '06)*, pages 401 –405, December 2006.

[57] J. Tian and L. Chen. Adaptive multi-focus Image Fusion Using a Wavelet-Based Statistical Sharpness Measure. *SignalProcessing*, 92(9):2137 – 2146, September 2012.

[58] R. Mariani. Microsoft. Garbage Collector Basics and Performance Hints, April 2003. https://msdn.microsoft.com/en-us/library/ms973837.aspx [07/01/2015].

[59] Microsoft. IDisposable Interface. https://msdn.microsoft.com/en-us/library/system.idisposable%28v=vs.110%29.aspx [07/01/2015].

[60] EMGU CV. Working with Images. http://www.emgu.com/wiki/index.php/Working_with_Images [07/01/2015].

[61] Microsoft. Parallel.For Method (Int32, Int32, Action<Int32>). https://msdn.microsoft.com/en-us/library/dd783539%28v=vs.110%29.aspx [07/07/2015].