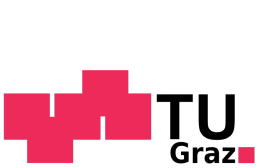



Graz University of Technology

Institute for Computer Graphics and Vision

Master's Thesis

Multi-Frame Film Restoration - Dirt and Dust Detection

Stefanie Wechtitsch

February 2012, Graz (Austria)

Master's Thesis Supervisors:

Univ.-Prof. Dipl.-Ing. Dr.techn. Horst Bischof

Dipl.-Ing Hannes Fassold

Abstract

Automatic film restoration tools are increasingly demanded by film archives due to the huge amount of data and due to the high cost of manual restoration. In this work we focus on automatic detection of single-frame-defects. Related work lacks in considering motion estimation errors and motion discontinuities, which result in high false alarm rates. The main issue resulting from false detections is the risk of introducing disturbing artifacts into the restored results. We propose a novel single-frame-defect detector which addresses this issue by considering pathological motion. First, a primary response measure is presented, which indicates the likelihood of each single pixel to be defected. Due to the probability based detection measure, essential information can be kept throughout the entire process. Second, the co-support operator is integrated, which effectively reduces false detections due to noise. Furthermore, the operator completes dust spots by adapting pixels based on their neighborhood. Finally, areas of pathological motion are obtained by analyzing the motion vector field. Based on deductions made from the motion field, a variety of measures are inferred which decreases the probability values of pixels in affected areas. A comparative evaluation has been performed and the results are illustrated using receiver operator characteristics (ROC) curves, where the novel approach is compared against several state of the art techniques. Experimental results show that the proposed technique outperforms all other techniques regarding accuracy and robustness to motion estimation issues.

Keywords: Film restoration, single-frame-defect detection, motion estimation, motion field analysis, pathological motion, occlusion and uncovering, co-support operator, noise suppression, ROC analysis.

Kurzfassung

Riesige Datenmengen und hohe Kosten für manuelle Restaurierung haben zu einer verstärkten Nachfrage an automatisierten Werkzeugen für die Filmrestaurierung geführt. In der vorliegenden Arbeit liegt der Fokus auf die automatische Detektion von Defekten in Einzelbildern. Vorgängerarbeiten waren vor allem mit Problemen in der Bewegungsschätzung und Bewegungsunstetigkeiten konfrontiert, aus denen üblicherweise eine hohe Rate an Falschdetektionen resultiert. Daher fehlte in vielen Ansätzen die ausreichende Berücksichtigung solcher Fehler der Bewegungschätzung, oder sie wurden überhaupt ignoriert. Das Hauptproblem von Falschdetektionen ist das erhöhte Risiko im restaurierten Ergebnis störende Artefakte wieder zu finden. Wir präsentieren einen neuartigen Detektionsansatz, der diesem Problem durch Berücksichtigung von pathologischer Bewegung entgegenwirkt. Zuerst wird ein Maß für die Primärdetektion vorgestellt, das für jedes einzelne Pixel die Wahrscheinlichkeit angibt, defekt zu sein. Aufgrund dieses Wahrscheinlichkeitsbasierten Detektionsmaßes kann über den ganzen Detektionsprozess hinweg auf wichtige Information jedes Bildpunktes zugegriffen werden. Weiters wird ein neuartiger Operator präsentiert, der die Fähigkeit hat, Falschdetektionen aufgrund von Rauschen effektiv zu eliminieren. Zusätzlich vervollständigt dieser Operator potentialle dust spots, indem er jedes Pixel basierend auf seiner Umgebung adaptiert. Zuletzt werden Regionen die pathologische Bewegung beinhalten durch eine ausführliche Analyse des Motionfeldes ermittelt. Aufgrund dieser Analyse werden verschiedene Indikatoren von pathologischer Bewegung abgeleitet, die die Wahrscheinlichkeitswerte für die Pixel in den betroffenen Arealen senkt. Der präsentierte Detektionsalgorithmus und ausgewählte state of the art Methoden wurden mittels ROC (Receiver Operator Characteristics) Kurven evaluiert und zusätzlich qualitativ analysiert. Die Ergebnisse dieser Experimente zeigen, dass dieser Detektionsalgorithmus alle anderen Methoden im Hinblick auf Präzision und Robustheit gegenüber Problemen in der Bewegungsschätzung übertrifft.

Acknowledgments

I wish to express my sincere gratitude to Prof. Dipl.-Ing. Dr. techn. Horst Bischof for his supervision, continuing support and constructive critics. I am deeply indebted to my supervisor Dipl.-Ing. Hannes Fassold whose encouragement, guidance and support from the initial to the final level enabled this master's thesis to be successfully completed. Thank you for always being available, for discussing about problems, for reviewing and for all your advice and patience.

Furthermore, I would like to thank Dipl.-Ing. Werner Haas and Dipl.-Ing. Georg Thallinger who made this cooperative work possible, Dipl.-Ing. Peter Schallauer for his support and valuable hints and all other people from JOANNEUM RESEARCH for being such good colleagues.

Grateful thanks to my sister, Mag. Simone Linordtner, as well as to my colleague, Dipl.-Ing. Werner Bailer, for offering their excellent editing skills in order to correct and improve this work regarding English style, grammar and content.

This work was funded by the Austrian Research Promotion Agency (FFG) under the Bridge project "Dust-NG II" (No. 827580). Thanks to everyone involved.

I owe my deepest gratitude to my family, in particular to my parents Mag. Franz and Christine Wechtitsch. Thank you for your love, for the continuing support and patience during all my time as student. Finally, I am heartily thankful to my boyfriend and all friends who supported me in any respect, for listening and for sharing many good times. My special thanks go to my roommate, Kerstin, who has been concurrently working on the final degree. Thank you for going with me through ups, downs and nights in the last months and for mutual motivation, distractions and laughs.

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am ……. ……..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

……. ……..
date (signature)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

According to UNESCO estimates, some 2.2 billion meters of 35mm films are currently stored at national and international film archives. This film collection consists of fundamental contributions to the world's cultural heritage, including for example records of history, politics, sports and entertainment [1].

All 35mm films recorded before 1950 were produced as nitrate films. In the Austrian film archive for example, the nitrate film collection represents the fundamental stock of Austria‘s contemporary and cultural history from 1896 into the 50s of the 20th century. Nitrate is a highly combustible and chemically volatile substance and threatened by continuing auto catalytic processes of decomposition [2]. Additionally, subsequent storage conditions, improper handling practices and even poorly maintained projectors have contributed to a continuing degradation of the material [1].

Hence, more than 100 years after the invention of film the audio-visual heritage of today‘s most important type of culture is imminent to disappear [2]. 90% of the movies shot in the silent era and half of all films produced before 1950 are already unrecoverably disrupted [3].

In order to stop the gradual disappearance, the storage conditions have to be improved and the remaining material needs to be preserved [4]. Nitrate films in particular are copied onto safety films and, more recently, onto polyester films, by

Video Standard	**Resolution**
SD	720 × 576
HD(720p)	1280 × 720
HD(1080p)(Full-HD)	1920 × 1080
2K	2048 × 1536
4K	4096 × 3072
8K	8192 × 6144

Table 1.1: Video standard resolutions, in which archived film material is typically scanned for digitalization and accordingly for the restoration.

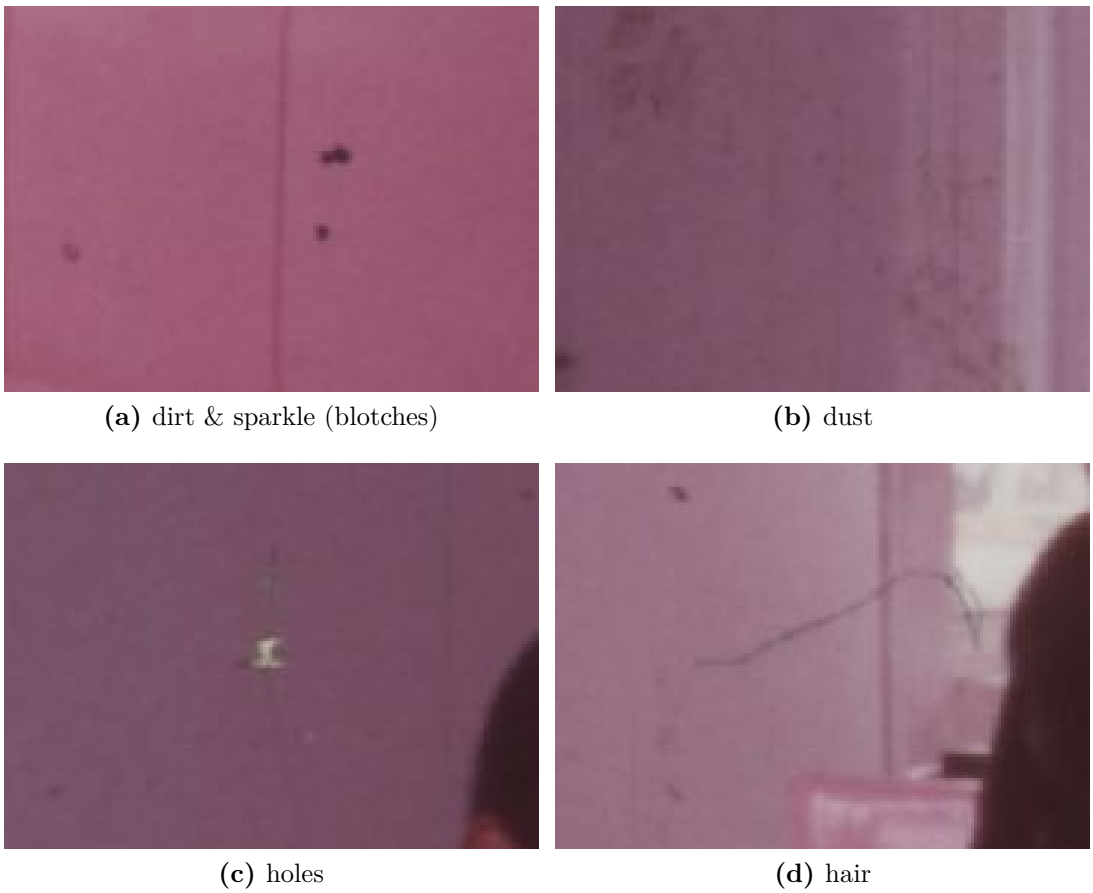
using specialist processing laboratories [2].

The environmental impacts, however, have left physical marks on the film strips [1], which would be simply recopied onto the new storage media. Hence the natural conclusion is to remove those defects before copying them to renewed storage, as they are impairing the viewing experience. By suppressing such impairments, the viewers aesthetic expectations can be met and the commercial value of the film and video sequences can be increased. Furthermore, it leads to more efficiency of video-coding algorithms [5] and since the originals are available in digital form, the films can be copied any number of times [1].

1.2 Digital film restoration

In traditional, analogue film restoration photochemical techniques are applied, which mostly avoid further decay. However, such techniques cannot remove many of the defects. Modern techniques, in contrast, are supposed to provide a complete digital manipulation and correction system. First, archived films are digitalized by a high resolution film scanner. Generally, the resolution of the scanned images range from SD to 8K (see Table (1.1)).

The digital film restoration process is a two step approach. In the first step digital images are analyzed by using modern techniques of digital image processing in order to improve their quality and to detect corrupted regions.



(a) dirt & sparkle (blotches) **(b)** dust

(c) holes **(d)** hair

Figure 1.1: Overview of typical single-frame-defects occurring in archived films.

This is followed by the generation of a binary mask, which indicates corrupted regions by set pixels. In the next step, corrupted pixels are reconstructed based on surrounding spatio–temporal information e.g. by spatio-temporal interpolation, while uncorrupted ones stay untouched [6]. Ideally, the resulting restored images are as close as possible to the original.

In this master's thesis, the focus lies on the first part of digital film restoration process, the detection of one-frame-defects, including dirt and sparkle (blotches), dust, cuts, tears, holes, scratches, hairs, etc., illustrated in Figure (1.1). They are the most frequent and annoying artifacts in archived films. Dependend on wheater

Characteristics	**utilized for spatial filtering**	**utilized for temporal filtering**
random distribution		×
high contrast to uncorrupted background		×
smooth interior	×	
high intensity difference to neighborhood	×	

Table 1.2: Common characteristics of single-frame-defects. Those characteristics are utilized by different single-frame-defect detectors in order to find as much as possible of present defects.

particles adhered on positive or negative film stock, they appear as dark or bright spots at different positions and have a varying size and shape.

One-frame-impairments have valuable characteristics, which can be exploited by detection algorithms. In the first place, blotches appear randomly distributed, so the assumption of a temporal discontinuity is applied, which means corrupted parts hardly ever appear at the same spatial location in consecutive frames. Generally, they have arbitrary shape, size, a varying range of intensity and opacity. Furthermore, blotches hardly are spatially consistent at their borders. Hence the intensities of the interior of a blotch are significantly different from those of its uncorrupted neighborhood. Finally, blotches are local artifacts and form coherent regions with almost the same brightness [7]. These common characteristics are summarized in Table (1.2).

Serious problems are the influence of local/global motion and the presence of other film degradations such as vibration, flickering or film grain noise at various levels. Therefore the challenge lies in the correct detection of corrupted regions even though the input images have varying characteristics. At the same time the aim is to keep the number of false alarms (FA), which represents uncorrupted pixels falsely detected as dirt, as low as possible. A further difficulty is the huge amount of data to be processed at high resolution (2K or higher) [8]. Due to the large amount of films to be restored, automated techniques are required.

This master's thesis was motivated by the need of a new state of the art single-frame-defect detection algorithm for DIAMANT, which is a well-established product in the marketplace of digital film restoration, developed by JOANNEUM RESEARCH[1]. The software already offers film restoration tasks in order to eliminate undesirable artifacts, such as removal of single-frame defects, images stabilization and correction of brightness and color instabilities. Among others, DIAMANT has already been used for the restoration of the classic TV detective series *Derrick* or the cult film *Metropolis*.

However, despite having a great variety of different approaches from literature available, to date, no satisfying solution exists for the problem of single-frame-defect detection and removal. In this master's thesis, the dirt and dust detection algorithm used in the DIAMANT film restoration tool should be improved. In particular, we want to focus on motion estimation (ME) problems in order to avoid misdiagnoses of defects and true data disruption caused by ME failures.

1.3 Outline of the thesis

This thesis is organized as follows: In the next chapter an overview of related work is presented. Related methods are divided in spatial, temporal and spatio-temporal single-frame-defect detection approaches, which are described and discussed in detail. Furthermore, previous work concerning various post processing mechanisms and the detection of pathological motion (PM) is addressed. The third chapter explains general assumptions and the main issues of motion estimation (ME) methods. In addition, a brief survey of four ME algorithms, which manifests different characteristics, is given. In the fourth chapter the proposed single-frame-defect detection algorithm is presented. We elaborately describe the developed components consisting of primary detection measure, an effective noise suppression method and several protection measures. These measures focus on the detection of precarious regions emerged from ME errors and are mainly inferred

[1] http://www.joanneum.at/

from analysis of motion vector fields. Experimental results are given in the fifth chapter, where the novel approach is evaluated against well known state of the art methods and the detector integrated in the DIAMANT film restoration software. A comparative, quantitative evaluation is performed by using the well known ROC approach. Additionally, we provide a qualitative performance analysis of all examined techniques followed by a detailed discussion of the obtained results. The last chapter gives a conclusion and an outlook for future work.

Chapter 2

Related Work

2.1 Introduction

In literature, detection algorithms commonly model the corruption of a frame as a binary mixture between the original, clean image and an opaque dust mask [9]. Thus the degradation of a clean image can be modeled as

$$I_n(z) = (1 - b(z))E_n(z) + b(z)c(z) + \mu(z) \tag{2.1}$$

where $E_n(z)$ and $I_n(z)$ denote the intensity at each site $z = (x, y)^T$ in the original and the degraded input frame n, respectively. $b(z)$ specifies a binary blotch mask that determines whether a defect is present, $b(z) = 1$, at pixel location z or absent, $b(z) = 0$. $c(z)$ is the pixel intensity in the corrupted frame and $\mu(z)$ indicates the additive Gaussian noise with variance σ^2 [10]. In this master's thesis, the model from Equation (2.1) will be slightly adapted and will be examined in Chapter (4).

2.2 Detection of single-frame-defects

The detection of missing data from the original frame E_n is mostly done by estimating $b(z)$ from Equation (2.1) at each pixel. Existing algorithms mainly use heuristics and generally try to exhibit the temporal and spatial discontinuity characteristic of single-frame-defects, already mentioned in Section (1.2) and listed in

(a) frame $(n-1)$ **(b)** frame (n) **(c)** frame $(n+1)$

Figure 2.1: Subset of 3 succeeding frames of image sequence *Dance2* (frame number 63 - 65).

Table (1.2). Hence existing work can be arranged in the following three categories:

1. Spatial filtering methods
2. Temporal filtering methods
3. Spatio-temporal filtering methods

2.2.1 Spatial filtering methods

Considering single-frame-defects as spatially impulsive events, they can be detected and recovered via filtering. Consequently, spatial filtering methods are often used as an alternative or, as later explained, as a complement to motion compensation. Prominent are those methods using median or morphological filtering [5].

In 1987 Nieminen et al. [11] proposed a multi-level median filter (MLF), which uses hierarchical median operations to reduce the influence of outliers while preserving edges [12]. Buisson et al. [13] used a top hat morphological filter, because of its ability to detect specific patterns such as dust and hair [14].

Two conventional methods of this group are standard spatial median filter (SSMF) [15] and Lower–Upper–Middle (LUM) filtering [16]. In both methods for each pixel $z = (x, y)^T$ in the current frame I_n, a window W of radius r is defined as

(a) Window size: 5x5 **(b)** Window size: 7x7 **(c)** Window size: 11x11

Figure 2.2: Detection results of LUM: Obtained dust mask by using a varying window size. For all shown results the threshold $\tau = 10$, $k = 5$ and $l = 10$.

$$W(x, y, r) = \{I(x_1, y_1)\},\ |x_1 - x| \leq r,\ |y_1 - y| \leq r,\ \forall x_1, y_1 \tag{2.2}$$

so that the total number of pixels in W is $N = (2r + 1)^2$ and $N_0 = \frac{(N+1)}{2}$. Furthermore, $W = z_{(1)}, z_{(2)}, \ldots z_{(N)}$ and the rank ordered set is then given by $z_{(1)} \leq z_{(2)} \leq \cdots \leq z_{(N)}$.

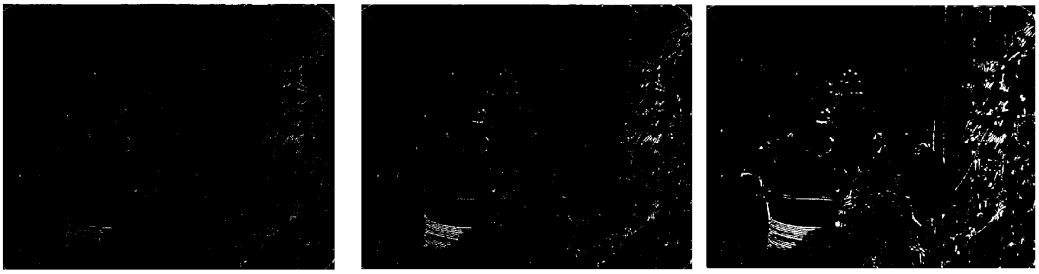
The central pixel in the original, current frame is denoted as z' and the corresponding central pixel in the filtered image as $z'_{filtered}$. In SSMF $z'_{filtered}$ is set to $z_{(N_0)}$.

In LUM the output is assigned to the upper or to the lower local median values, defined in Equation (2.4). Two parameters, k and l, are introduced for smoothing and sharpening, where $1 \leq k \leq l \leq N_0$. The corresponding outputs of the smoothing and sharpening processes are given by

$$z_L = \text{median}\{z_{(k)}, z', z_{(l)}\} \tag{2.3a}$$

$$z_U = \text{median}\{z_{(N-k+1)}, z', z_{(N-l+1)}\}, \tag{2.3b}$$

the filtered output is then specified as



(a) Window size: 5x5 **(b)** Window size: 7x7 **(c)** Window size: 11x11

Figure 2.3: Detection result of SSMF: Obtained dust mask by using a varying window size. For all shown results the threshold τ is set to 10.

$$z'_{filtered} = \begin{cases} z_L, & \text{if } z' \leq \frac{(z_L+z_U)}{2} \\ z_U, & \text{otherwise.} \end{cases} \tag{2.4}$$

If $I_{filtered,n}$ is the output image after filtering and τ is a predefined threshold, the dirt mask b can be determined as

$$b(z) = \begin{cases} 1, & \text{if } |I_{filtered,n}(z) - I_n(z)| > \tau \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

In both, SSMF and LUM, the detection results are very sensitive to the selected size and shape of W. Although it was reported in [5] that SSMF has a better detection performance than LUM, the resulting false detections are of unacceptable degree. FA particularly occur close to sharp edges or in textured regions [14].

In Figure (2.2) the detection results of the LUM method are illustrated. For all shown results the threshold τ is chosen to be 10, and the smoothing and sharpening parameters k and l are set to 5 and 10, respectively. In Figure (2.3) the results of the SSMF method are shown, where threshold τ was chosen to be 10 as well.

2.2.2 Temporal filtering methods

Methods belonging to this category are based on the hypothesis that single-frame-defects are temporally impulsive and thus randomly distributed over frames, hence they can be detected by inter frame processing [17]. In most cases 3 frames are considered.

The overall first contribution to the electronic detection and concealment of film dirt was done by Storey [18] as early as 1983. The idea of his hardware-based system was to flag a pixel as dirt if its corresponding absolute pixel differences between the current frame and each of its neighbors were high. However, the moving objects in the scene were not taken into account, thus it cannot correctly separate moving areas from blotched regions [19].

An expansion of this idea was presented by Kokaram [20], who extended Story's work by introducing motion compensation. He presented the Spike detection index (SDIa), which allows the detection of temporal discontinuities between the current frame I_n and the motion compensated preceding C_{n-} and subsequent C_{n+} frames, under the assumption that those neighboring frames are uncorrupted. The two difference images D_{n-} and D_{n+} are obtained using

$$\begin{aligned} D_{n-}(z) &= I_n(z) - C_{n-}(z), \\ D_{n+}(z) &= I_n(z) - C_{n+}(z) \end{aligned} \tag{2.6}$$

at each pixel $z = (x, y)^T$. The dust mask b_{SDI_a} for the SDIa-detector results from computing the binary value at each pixel z

$$b_{SDI_a}(z) = \begin{cases} 1, & \text{if } D_{n-}(z) > \tau \text{ and } D_{n+}(z) > \tau \\ 0, & \text{otherwise,} \end{cases} \tag{2.7}$$

where τ is a predefined threshold, which measures the discontinuity of the two frames, and $b_{SDIa}(z)$ is a detection field variable set to 1 at sites that are corrupted.

The SDIp method [20], which is a further extension to SDIa, additionally implies the constraint that corruption does not occur in identical locations in consecutive frames, hence, it requires sign consensus of the two frame differences

D_{n-}, D_{n-}. Thus, a pixel in SDIp is declared as dirt if it satisfies Equation (2.8).

$$b_{SDIp}(z) = \begin{cases} 1, & \text{if } \begin{array}{l} |D_{n-}(z)| > \tau \text{ and } |D_{n+}(z)| > \tau \\ \quad \text{and } D_{n-}D_{n+} > 0 \end{array} \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

Schallauer et al. [21] extended the SDIa method in a similar way. They even strengthened the constraint that the two motion compensated neighboring frames have to be similar as well. Hence a pixel is taken as dirt if both, its absolute differences between current frame and the two motion compensated images, exceed a first (higher) threshold, as already used in SDIa and SDIp, while at the same time the absolute difference between the two motion compensated neighbors is less than a second (lower) threshold defined in Equation (2.9). Consequently, this method is often denoted as a double-threshold median filter (DTMF) [22]. As this method was developed at JOANNEUM RESEARCH it is here referred to as SDIjrs. Note that this method is the basic detection algorithm which is used in the DIAMANT film restoration tool [21].

The dust mask is obtained by using the following equation

$$b_{SDIjrs}(z) = \begin{cases} 1, & \text{if } \begin{array}{l} |D_{n-}(z)| > \tau_1 \text{ and } |D_{n+}(z)| > \tau_1 \\ \quad \text{and } |C_{n-}(z) - C_{n+}(z)| < \tau_2 \end{array} \\ 0, & \text{otherwise} \end{cases} \tag{2.9}$$

where τ_1 and τ_2 are two given thresholds with $\tau_1 > \tau_2$.

Temporal median filtering is a common approach in dirt detection using the current frame and mostly the previous and the next neighbor, either motion compensated or not. If no motion compensation is included, the result is commonly worse, as a lot of FA occur in areas of moving objects. However, by using the additional constraint, the SDIp detection algorithm is able to achieve a significant improvement in detection performance when compared with the results of the SDIa detector [10]. Although different strategies are used to filter FA, SDIp and SDIjrs produce similar results. SDIp, however, seems a better solution than SDIjrs as it

(a) SDIa **(b)** SDIp **(c)** SDIjrs

Figure 2.4: Detection results of the SDIx methods. For all algorithms a threshold of $\tau = 10$ was used, for the SDIjrs the thresholds are set to $\vec{\tau} = (10, 7)$.

is more controllable due to its single threshold [5]. In Figure (2.4) the detection results of each of the SDIx methods are illustrated. For each of the shown results we have used motion compensated neighboring frames, by applying TV-L1 optical flow (see Section (3.2.1) for details).

2.2.3 Spatio-temporal filtering methods

Spatio-temporal methods extend spatial filtering to the temporal domain thus they combine useful features from both groups. Several attempts have been made, to simply extend the spatial filtering methods to three consecutive frames. Arce et al. [23] introduced a multi-stage order statistic (MOS) filter as an extension to the min/max MLF [11]. Four groups of sub-windows in three non-motion-compensated frames are used. MOS filtering has proven very useful in image processing, as well for image sequence noise suppression, because of the additional temporal information [12]. At the same time, Arce [23] also proposed a three-frame LUM variant, also denoted as LUM', for image smoothing purposes, where the filter was further applied to a 3 x 3 x 3 spatial-temporal window. MOS and LUM' methods were effectively designed for spatio-temporal filtering, although the performance of both is sensitive to the designed shape of sub-windows. LUM' seems superior to MOS due to the fact that it takes more pixels from neighboring frames in filtering. However, this may also cause more FA. In general, LUM' performs better only in the absence of moving edges in the images [5]. In the

same manner, Alp et al. presented another multi-level median filter, the ML3D filter [24], which utilizes two-stage median filtering in three consecutive frames [5]. For this algorithm two groups of sub-windows are defined in three frames and their median values are calculated. The output is obtained by taking the median of those two values and the current pixel value. This filter is applicable for noise removal in image sequences as well [25]. A. Kokaram [20] improved the ML3D filter by extending the method to 5 groups of sub-windows and motion compensated neighbors, which is denoted as ML3Dex. According to Kokaram, the ML3D filter provides better impulse noise rejection than the LUM' method proposed by Arce. Hamid et al. [26] proposed soft morphological filtering (SMF) in three non-motion-compensated frames. They used a genetic algorithm to learn the size and shape in the filters, supervised by corrupted and uncorrupted sequences [14]. A detailed description can be found in [25]. SMF was introduced to prevent the high FA rates at edges, caused by median and morphological filters. It results in less FA than LUM' and ML3Dex for fast moving objects. However, SMF is impractical for most film restoration applications. Firstly, the learning step is very slow as the filter parameters have to be adapted to each new sequence [14]. Secondly, it needs a sufficient number of representative dirt samples for training purposes in order to optimize the filters [25]. In 1996, the rank-ordered difference (ROD) detection method was presented by Nadenau et al. [27]. The algorithm works by measuring the "outlierness" of the current pixel in comparison to a set of others. A spatio-temporal window is used to arrange 6 pixels from neighboring (three from previous and three from next) frames in a ranked order and then a three-stage thresholding strategy is applied. The choice of the shape of the region from which the other pixels are chosen is arbitrary. Let $I_n(z)$ denote the intensity of a pixel at a spatial location $z = (x, y)^T$ in frame n. Let $R_{n;i}(z)$ form a set of $i = 6$ reference pixels, obtained from spatially co-sited pixels and their vertical neighbors in motion compensated previous and next frames. The elements of the set $R_{n;i}(z)$ are ranked in a list $[r_1, r_2, ..., r_6]$ so that r_6 is the maximum element. Then the median is extracted as $m_r = \frac{(r_3+r_4)}{2}$. Three thresholds need to be defined so that $\tau_1 \leq \tau_2 \leq \tau_3$. Furthermore the e_k values are obtained by

$$e_k = \begin{cases} I_n(z) - r_{(7-k)}, & \text{if } I_n(z) > m_r \\ r_k - I_n(z), & \text{otherwise} \end{cases}$$

where $k \in 1, 2, 3$. A pixel is then detected as dirt if Equation (2.10) is satisfied:

$$b_{ROD}(z) = \begin{cases} 1, & \text{if } e_1 > \tau_1 \text{ or } e_2 > \tau_2 \text{ or } e_3 > \tau_3 \\ 0, & \text{otherwise.} \end{cases} \tag{2.10}$$

ROD is generally more robust to ME errors, it achieves higher correct detection rates and lower FA than the other methods mentioned above [28]. The three thresholds control the number of correct detections and FA. However, the difficulty of defining these thresholds constrains the effectiveness of the ROD method [29].

Thus a simplified ROD detector (sROD) which uses only one threshold has been proposed by Roosmalen [30]. By letting $\tau_2, \tau_3 \to \infty$ the output of the sROD detector is completely determined by τ_1. The output $d_n(z)$ is given as:

$$d_n(z) = \begin{cases} min(R_{n;i}(z)) - I_n(z), & \text{if } min(R_{n;i}(z)) - I_n(z) > 0 \\ I_n(z) - max(R_{n;i}(z)), & \text{if } I_n(z) - max(R_{n;i}(z)) > 0 \\ 0, & \text{otherwise .} \end{cases} \tag{2.11}$$

A blotch is detected if the intensity of the current pixel lies far enough outside that range. What is considered "far enough" is determined by τ_1:

$$b_{sROD}(z) = \begin{cases} 1, & \text{if } d_n(z) > \tau_1 \\ 0, & \text{otherwise} \end{cases} \tag{2.12}$$

with $\tau_1 \geq 0$.

The sROD detector takes local motion changes into consideration and has an improved blotch detection performance over the SDIx detectors [29]. Basically, it compares the current pixel to the range of values, determined by the minimum and maximum of the set, using the same set of pixels as the ROD method [14]. The simplified version does not order the reference pixel by rank, hence it is com-

(a) ROD, $\vec{\tau} = (11, 13, 15)$ **(b)** sROD, $\tau = 10$

Figure 2.5: Detection results of the methods ROD and sROD.

putationally more efficient. The drawback of the sROD detector is the higher false alarm rate (FAR) when choosing a smaller threshold in order to achieve a higher correct detection rate (CDR) [28].

Gangal et al. [19] extended the ROD method to five consecutive frames to improve the detection accuracy in heavily corrupted, occluded and uncovered areas.

Combined Methods

Buisson et al. [13] presented a combination of SDIa and a morphological detector which uses spatial properties of deteriorations. In contrast, Ren and Vlachos [12], [31] tried to disclaim motion compensation. Based on a combined method they incorporated confidence weighting to avoid information loss. They argued, that motion-based approaches often fail in areas where the motion cannot be accurately estimated. This approach has a low complexity due to the lack of motion compensation, but for the same reason it seems to be only applicable to sequences with limited local motions [14]. However, in [5], they extended their method by additional global motion compensation. A better modeling of the combination of temporal and spatial detection is proposed by Tilie et al. [14]. Temporal detection is used to exclude FA in the spatial domain due to the ambiguity between dirt clusters and objects of a similar spatial structure. Spatial detection is used to either confirm the detections from the first step or to discard them if they were falsely detected e.g. due to failing ME. The presented fusion method also promises

lower computational costs.

In order to increase the ratio between correct detections and FA by using the sROD method Wei et al. [28] introduced adaptive multi-thresholding. Their detector uses the most appropriate threshold for each blotch by convergence confinement. To avoid deviations caused by inaccurate motion vectors also texture matching is applied.

In [32] it is shown that the performance of blotch detectors can be improved significantly by taking statistical influence of noise on the detection mechanism into account. Further improvements are achieved first by using a double-stage detection strategy and second by a constrained dilation technique.

Model based approaches

According to [10] a common problem with approaches to dirt and dust detection is, that they always address single issues separately. Statistical methods, e.g. a Bayesian framework using a Markov random field (MRF) prior, are able to model context dependent entities such as image pixel intensities and other spatially correlated features. Thus the Bayesian approach has been evolved into a unifying framework that treats motion, defects and noise jointly [20], [33], [34]. To resolve the problem of combining different sorts of prior information about common frame defects, also auto-regressive (AR) [35] models have been used.

In the MRF based approach proposed in [36], the detection of the blotches in image sequences is formulated as a maximum a posteriori (MAP) estimation problem. They incorporated a moving-edge detector to address the problem of incorrect detection due to poor motion compensation at the moving edges [36]. By separating the blotch-edges from the moving edges the degree of FA in the detection can be effectively reduced.

Although these methods perform well in real situations, the main problem with model based approaches are the resulting mis-detections in cases of motion discontinuity [10], such as occlusion and the high computational cost, which becomes intractable when the neighborhood order exceeds first or second order [14]. Kokaram et al. [37] reported, that model-based approaches using MRF and AR yield minor improvements compared to SDIa. This was confirmed by the results in [5], where

MRF methods perform equal or worse than SDIp in all tested image sequences. In [38], Nam et al. presented an improved MRF model with lower computational cost and higher blotch detection performance than the existing MRF-based methods. They promised a fast and efficient method by using diamond search (DS) based on a new matching criterion, a modified cost function and a moving edge detector.

Reduction of false alarms (FA) by post processing

High detection rates are desirable, but only useful if the number of false alarms (FA) is acceptable as well. Falsely detected pixels may produce visible artifacts in the corrected image frame. Several techniques have been applied to the restoration process, in order to examine this problem and to remove FA from the preliminary detected blotch candidates.

FA due to noise are addressed in [6], where a scale-space based noise reduction filter is presented. The filter is configured to get an optimal separation between signal and noise. Furthermore, three variants of post-processing to improve the general detection performance are proposed. They manipulate candidate blotches as objects rather than individual pixels. First, they compute the probability that the detector gives a specific response due to noise, which offers to compute the probability that a blotch of given size is wrongly detected. Depending on the given probability, the size and the corresponding detector response for each candidate blotch, it is decided whether to keep it in the detection mask or not. The second method completes blotches which are only partially detected by applying a technique called hysteresis thresholding. Blotches are detected in two stages, by using a varying threshold. Then candidate blotches are preserved if corresponding candidate blotches in the other set exist as well, otherwise they are discarded. The last method is an iterative, constrained dilation technique, which fills small holes in candidate blotches that were missed by the detector. If a pixel's neighbor is flagged as being corrupted and its intensity difference to the certain neighbor is small the pixel will also be flagged as defected.

Similarly, morphological post processing, such as a constrained dilation algorithm, has been proposed in [39] in order to reduce FA which usually arise from

edges. Furthermore, a combination of SDIa and autoregressive (AR) method for the detection of rather large blotch regions is presented. This post-processing improves the detection rate of SDIa but FA caused by local object motion and incorrect global motion compensation are not eliminated adequately [29].

Instead of a pure pixel-based blotch detector, the OFST technique [40] uses an object-detector index (ODI) to detect temporal discontinuities, as the methods described in Section (2.2.2) do. Statistical object features such as the internal contrast, i.e. the difference between minimal and maximal intensities, and the difference of internal and external mean values are computed to further identify the blotch candidates.

In [8] and later in [41] a two-step FA reduction method is proposed. They use a semi-automatic neuronal network based detection followed by pixel- and object-based post-processing. The pixel-based algorithm compensates motion in order to decrease FA, while the object based method further classifies each detected blotch by machine learning techniques using image features.

A new post processing method based on region labeling is presented in [42], where noise is removed while the connectedness of blotches can be preserved. They also apply a motion vector correction method based on the blotch mask for improving the result.

A novel pixel-based correction method, presented by Güllü et al. [29], iteratively calculates correction priorities for each pixel on the contour of the blotch mask. It also corrects pixels with high priorities, updates the blotch mask and recalculates priorities until the blotch mask is either empty or nothing changes any more.

Spatio-temporal methods achieve a better performance than spatial or temporal methods alone. Derived from recent work, the detection performance of spatial, temporal and spatio-temporal methods has to be improved by post processing methods in order to achieve acceptable results. Although, additional methods can improve the detection results, they often lead to a higher computational cost. Generally, FA due to noise can be significantly reduced, when previously mentioned post processing techniques are applied. Techniques, in order to correct incorrectly estimated motion vectors have already been proposed, such as in [42], but the

correction is mainly limited to motion vectors in blotched regions. Incorrectly computed motion vectors as a result of complex motion, motion discontinuities (occlusion and uncovering), etc. cannot be repaired. Thus FA due to so called PM persist [14].

2.3 Detection of Pathological Motion (PM)

The term "Pathological Motion" (PM) subsumes all kinds of object motion that cannot be easily modeled by current motion estimation (ME) algorithms. Some common examples for PM are fast and irregular motion, occlusion and uncovering, motion blur, strong zooming, etc. A taxonomy of PM is presented in [43].

For dealing with PM existing approaches can largely be categorized in two classes [44]. The first one rather detects regions which are likely to be part of PM. In most cases a segmentation of foreground and background is performed in order to find motion discontinuities, which indicate occlusion and uncovering areas. Then blotch treatment is adapted appropriately in these areas, as it is done e.g. in the work of Kent et al. [45]. In the second class color statistics of PM are used. The facts, that blotches typically appear as regions consisting of similar intensity values and that motion blur is the smearing of an object in the foreground over the background are effectively utilized in [46]. Complex events resulted from PM and those caused by image artifacts are classified based on inter-frame segment matching.

To prevent mis-diagnosis of missing data due to PM, Corrigan et al. [44] also propose the consideration of more temporal information such as 5 frames. They estimate four temporal discontinuity fields as binary masks by exploiting the local smoothness assumption of motion fields. The violation of the local smoothness assumption is indicated by high divergence values in the motion field. Independent of the blotch detection process, the problem of the detection of motion discontinuities was addressed already in related work. In [47] the motion is computed in the forward and the reverse direction in order to find occlusion areas. The obtained flow fields are constrained to be compatible, thus these areas where the compatibility is not achieved, indicate occlusions. A symmetric stereo matching model to handle occlusion is proposed by Sun et al. [48]. They utilize the visibility con-

straint, which enforces consistency between occlusion in one image and disparity in the other. It is a iterative optimization algorithm, by alternating updating the disparity map and the occlusion map.

A motion segmentation method including an occlusion detector is developed in [49] which is based on differential properties, such as corners and edges, of the spatio-temporal domain. These structures are detected by a Harris corner detector in order to extract motion boundaries.

Probabilistic approaches were presented as well, such as the work of Lim et al. [50], in which the ME error is integrated within a bi-directional Bayesian framework. In [51] an iterative occlusion aware refinement step was introduced jointly to the ME. By using the mapping uniqueness criterion, occluded pixels are detected, if more than one reference pixels mapping to the same position when following the motion vectors of one frame to the subsequent one.

2.4 Discussion

As shown in this chapter, several attempts have been made to address the problem of single-frame-defect detection. Spatial median [16] and morphological filters [13] have been extensively used because of their ability to eliminate outliers while preserving edges. They result in a low computational cost, as they do not include temporal information. However, these spatial detectors fail to detect low-contrast or exceeding filter size botches. It may generate many FA on sharp and textured regions [17], which may lead to heavy degradation of visually significant edges [52]. In most cases the resulting images of spatial filtering are of unacceptable quality [12].

Temporal median filtering, on the other hand, is a common approach complementary to bi-directional motion compensation. Commonly, the current frame and its two (motion-compensated) neighbors are used. Also, more temporal support is possible, as shown in [19], where the proposed detection method provides an increased robustness to PM, noise and blotched areas, but it produces more FA along moving edges. Consequently, increasing the temporal aperture does increase the computational cost, while it does not necessarily increase the detection performance [30]. Although temporal filtering methods are capable of achieving a

high CDR they commonly result in too many FA [39], in particular at presence of noise, textured background or complex motion. Thus the detection performance in temporal filtering methods is very sensitive to the used ME algorithm.

Spatio-temporal methods achieve better performance than spatial or temporal methods alone. Although they consider both the spatial and temporal discontinuity characteristics of blotches, they are likely to produce FA in presence of noise or complex motion as well. According to a comparative evaluation in [5], ROD resulted in the highest accuracy while its computational cost was lower than that of MRF but higher than SDIp. The ROD detector, however, requires the setting of three thresholds, which are hard to define. In recent work, [6], [8] the sROD detection technique has turned out to be an appropriate method for the film restoration process. The sROD detector takes local motion changes into consideration [29] and, moreover, it involves only one threshold. The sROD detection method has the drawback of producing more FA when reducing the threshold in order to achieve a higher correct detection rate (CDR) [28]. Thus, it seems to be necessary to apply a post processing step, e.g. an adequate method presented in Section (2.2.3).

The incorporation of confidence weighting [12] offers a possibility to measure the likelihood of corruption for each pixel. Having probability values available instead of binary values is advantageous, since the final decision by applying a threshold can be postponed until post processing was performed.

Model-based schemes are able to consider how the characteristic of blotches can be incorporated in the problem definition itself [10]. However, since definitive statistical models are difficult to obtain, such methods have occasional constraints and will fail if the underlying statistical modeling assumptions cannot be satisfied or if accurate and robust motion compensation cannot be achieved [5]. So far, pixel based detectors have shown to achieve similar detection results as object based detectors at a fraction of the computational cost [6], which are intolerable in the domain of automatic film restoration.

Generally, the performance of methods that use motion compensation is better than of those which do neither motion compensation, nor any kind of false alarm reduction [5]. Non-motion-compensated approaches cannot easily distinguish between genuine dirt clusters and moving objects of a similar spatial structure and

also fail when such clusters exceed the filter size. Motion-compensated approaches on the other hand yield a worse performance when motion cannot be precisely estimated [12]. Thus, the motion estimate must be robust to the presence of dust spots and motion discontinuities, and the detector should work in tandem with the interpolation process [10]. Usually the motion compensated neighbors are used in the correction stage as well. Therefore, PM or noise falsely detected as blotches are a particular problem in restoration systems, because they possibly generate new visible artifacts that are more disturbing than the dust spots themselves [17]. Having areas of PM available, the detection of one-frame-defects can be adapted in this areas or even rejected.

Chapter 3

Motion Estimation

3.1 Introduction

Motion estimation (ME) is a vital component in many digital video processing algorithms, so it is in digital film restoration applications. The main objective of ME algorithms is to automatically estimate the transformation of one frame to the neighboring one within an image sequence. In local ME the transformation of one frame to another is represented as a dense field of two-dimensional displacement vectors, with components u and v for horizontal and vertical direction respectively, indicating the motion for each single pixel [53]. A motion field is obtained, which projects a 3D scene of moving objects (rotation, translation) onto a 2D image plane [54]. In this way, the motion of every individual object can be measured [55]. This dense field of motion vectors, also denoted as optical flow [53], gives essential information about relative motion of objects and the movement of the camera [56].

3.1.1 The translational model

Commonly three frames are considered in order to find single-frame-defects. Thus, motion compensated neighboring frames are computed by applying bi-directional motion compensation between the actual frame n and each of its two neighbors, $(n + 1)$ and $(n - 1)$. The same was applied to the temporal filtering methods described in Section (2.2.2). In general, ME algorithms assume that each im-

age can be constructed by displacing the pixels from the previous neighboring frame [34]. Consequently, Kokaram et al. [10] proposed the translational model. As bi-directional motion compensation is applied, the model is adopted for both directions as follows

$$\begin{aligned} I_n(z) &= I_{n-1}(z + d_{n,n-1}(z)) + e(z) \\ I_n(z) &= I_{n+1}(z + d_{n,n+1}(z)) + e(z) \end{aligned} \tag{3.1}$$

where I_n is the intensity function for the frame n at every pixel $z = (x, y)^T$, $d_{n,n-1}(z)$ and $d_{n,n+1}(z)$ are the two-component motion vector mappings of pixel z between the current frame n and the neighboring frames $(n-1)$ and $(n+1)$ respectively, and $e(z)$ is the Gaussian distributed error of the model [55].

3.1.2 Motion estimation constraints

In state of the art algorithms, dense optical flow is dominated by global variational models based on the work of Horn and Schunck [56], where they came up with a variational formulation of the optical flow problem.

$$\min_u \{ \int_\Omega |\nabla u|^2 + |\nabla v|^2 d\Omega + \lambda \int_\Omega (I_n(x + u(x)) - I_{n-1}(x))^2 d\Omega \} \tag{3.2}$$

I_{n-1} and I_n are the two images, between which the motion should be estimated, $u = (u(x), v(x))^T$ is a 2D displacement field, $x = (x, y)^T$ denotes the spatial image coordinated and λ defines a free parameter.

State of the art algorithms are designed in order to fulfill two constraints, namely the brightness and the smoothness constraint. The first one demands, that the actual pixel has a similar intensity value as the one, the motion vector points to. Secondly, neighboring points within a frame in general have similar velocities [56], thus in a neighborhood, the motion field has to be smooth. This observation is also known as the *Smooth Local Flow Assumption* [55].

In order to obtain a smooth motion field the regularization term (first term) of Equation (3.2) penalizes high variations in u. The optical flow constraint or the data term (second term) assumes, that the intensity values of $I_0(x)$ are similar to $I_1(x + u(x))$ [57].

3.1.3 Challenges for motion estimation

The accurate optical flow estimation is a challenging problem in computer vision as images commonly contain severe occlusion and non-rigid motion [58]. ME failures occur when the estimated motion vector does not point to the true displacement of a pixel [55]. They are likely to be caused in one of the following settings:

1. Presence of artifacts: As artifacts are impulsive defects, the containing pixels obviously cannot be matched within two consecutive frames and will result in failures. In particular large blotches, especially those exceeding the block size of the motion estimator, significantly impair the performance of the ME algorithm [55].

2. Presence of certain spatial textures: In regions of low texture, e.g. a cloudless sky, discriminative image features are absent in both the current and reference frame, thus an unambiguous match cannot be found [53]. Another well-known example of so called ill-conditioned image data [20] is represented by the aperture effect. This effect arises in textures showing only a single orientation or in periodic patterns. Even though such areas are textured, the motion direction is ambiguous and matching is difficult as the problem does not have a unique solution [53], [55].

3. Presence of PM: The phenomenon of PM has already been explained in Section 2.3. Problems emerge when a pixel's matching point is occluded in the reference frame. Such kind of occlusions often occur at motion discontinuities. Consequently, the precise identification of motion boundaries is specifically challenging in those areas [53].

Other problems appear when the ME method is supposed to deal with large motion, as usually, large motion vectors cannot be matched by ME methods or they are avoided due to the smoothness constraint. Another difficulty arises with scenes including illumination changes. In this situation the ME method is forced to violate the brightness constancy assumption.

3.2 Motion estimation algorithms

For the single-frame-defect-detection approach presented in this master's thesis, different ME algorithms are used, which are described in the following subsections. Although the problem of ME is not the subject of this work, an evaluation of mainly two algorithms is included, in order to compare restored results, when applying different ME methods. An overview on optical flow methods is given e.g. in [59] and [60]. All four algorithms were tested on the same machine in terms of computational cost. An eight-core Xeon processor, 2.67 GHz, 4GB RAM and a NVIDIA GeForce GTX 460 on a 64 bit Windows 7 Professional system was used. For each method we measured the average time for computing the motion of one image taken from a 2K (2048×1536) image sequence. The obtained motion results computed from the 3-frame-sequence shown in Figure (2.1) are visualized in an 8 bit gray scale image, so that positive translations, either in x or y direction, are coded in gray values ≥ 128 and negative translations in gray values < 128. In Figure (3.1 - 3.4) the estimated motion fields of each method are illustrated, where the upper parts of the two images show the translation in x and the under parts in y direction.

3.2.1 GPU based TV-L1 optical flow

The total variation (TV)-L1 optical flow algorithm [57], [61] developed at the Institute for Computer Graphics and Vision[1] was chosen in the first place, as variational methods to date represent the most accurate optical flow estimation. The algorithm is based on the TV-L1 energy functional, similarly to the method of Horn and Schunck [56], and uses the classical optical flow constraints already presented earlier in this chapter. Variational optimization is used to minimize a TV-L1 energy functional that assumes the fulfillment of these constraints.

In [57] the TV-L1 energy functional is given as

$$\int_{\Omega} \{\lambda\phi(I_0(x) - I_1(x + u(x))) + \psi(u, \nabla u|, \dots)\}dx, \tag{3.3}$$

where $\phi(I_0(x) - I_1(x + u(x)))$ is the image date fidelity, $\psi(u, \nabla u|, \dots)$ is the regu-

[1] Graz University of Technology (Austria), www.icg.tugraz.at

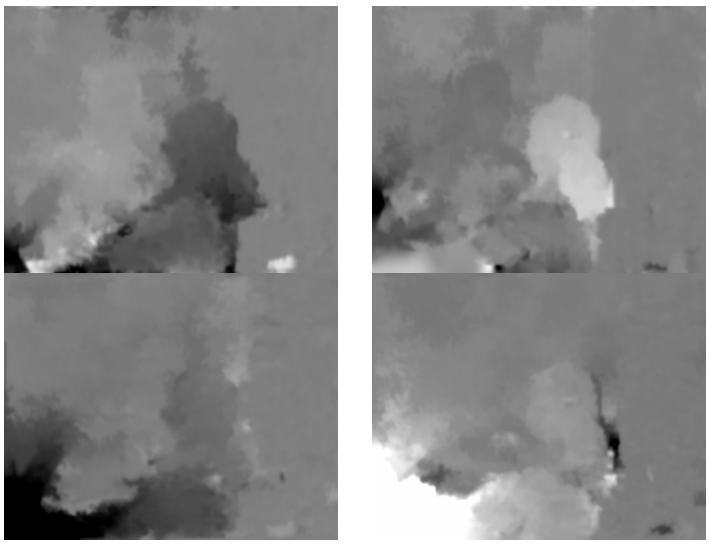
(a) center → previous **(b)** center → next

Figure 3.1: Motion fields computed by GPU-based TV-L1 optical flow. On the left, the motion field, computed from the center frame of previously shown subset of sequence *Dance2* to the previous one and on the right, the forward motion to the next frame is visualized.

larization term and λ weights between data fidelity and regularization force. The functional is based on the robust L^1 norm in the data fidelity term and a dual formulation of the TV energy. The TV regularization is image-driven and considers the strength and the direction of underlying image edges [61]. This algorithm is able to preserve discontinuities, can deal with illumination changes and occlusions, and offers an increased robustness against noise. Additionally, it can be effectively accelerated by modern graphics processing units (GPUs) [57]. The resulting visualized motion can be seen in Figure (3.1). The runtime of the GPU implementation for one 2K frame is 594,8 ms.

3.2.2 Pixel motion

The Pixel motion is used in the DIAMANT restoration tool. As there is no source code available, detailed information to the algorithm is not accessible. However,



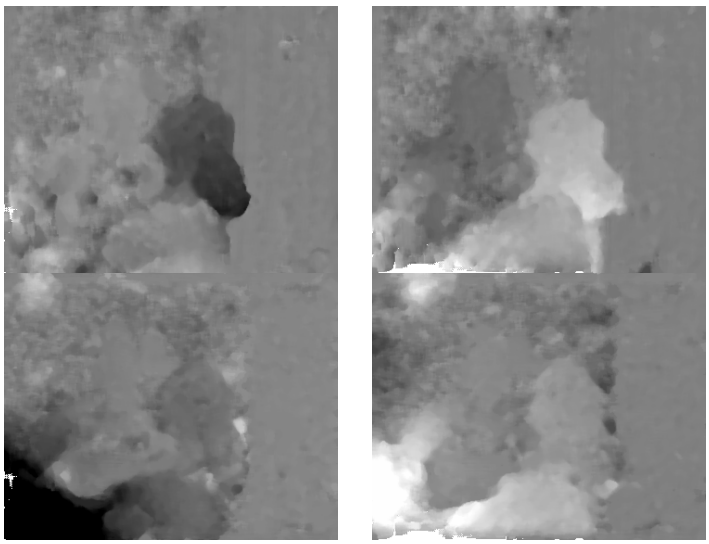
(a) center → previous **(b)** center → next

Figure 3.2: Motion fields computed by Pixel motion. The left visualization shows the motion of the chosen subset of frames from image sequence *Dance2* in previous direction and the right one in forward direction.

in Figure (3.2) the visualized motion of this algorithm is presented. The average computation time of one 2K frame for the CPU implementation of this algorithm is 2.57 sec.

3.2.3 HFVM

The HFVM method [62] was developed at JOANNEUM RESEARCH and can also be applied to the presented single-frame-defect detection method. The HFVM method is a feature based matching technique. Appropriate features for this approach are gray-level statistics, such as mean filters or variance measures, and other more sophisticated ones, like local frequency, edge gradient, Fourier-features, etc. Using a certain selection of these, a set of feature images is obtained by applying each feature to both images, the current frame and the reference one. A feature is derived from local properties of a pixel including its surrounding, defined by a search window. By using a variety of local image features, possible disadvantages



(a) center → previous **(b)** center → next

Figure 3.3: Motion fields computed by HFVM method. On the left the motion in the backward direction is demonstrated, and on the right, the motion in forward direction, applied to the same subset of image frames as the previously demonstrated algorithms.

of single features are equalized. For correspondence matching, a feature vector is obtained from the feature images for each pixel.

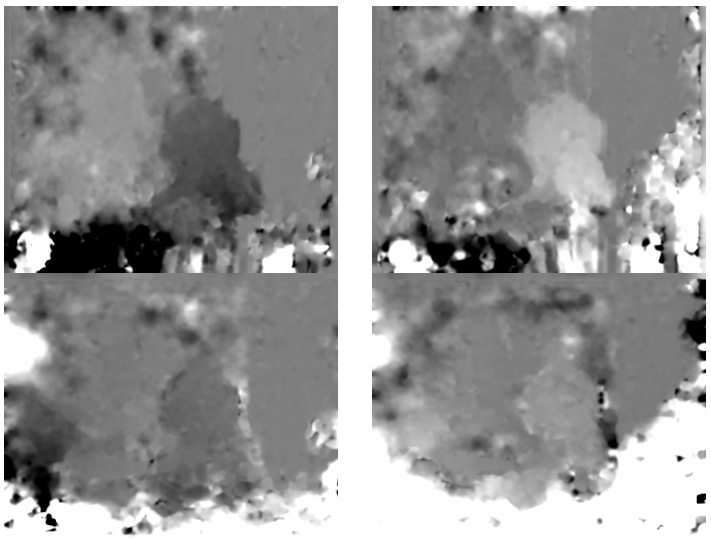
Then, feature vectors of the left and right image are compared to get homologous points: Once the feature images are created, a feature vector $\vec{f}(i,j)$ exists at each pixel location (i,j). The feature vectors $\vec{r}(i,j)$ in the reference frame are created in the same way. To find the best correspondence the Euclidean distance metric is used. In a final step errors are removed and remaining undefined disparities are interpolated. Paar et al. [62] use a well-known pyramid-based data structure for each of the images, namely a $3 \times 3/4$ Gaussian pyramid. 3×3 denotes the filter window and 4 is the reduction factor of one level to the next one. The number of levels depend on the maximum expected disparity. The method is applied to all images of the pyramid, beginning at the top level. In order to remove salt and pepper noise, median filtering is carried out on the disparity map. In addition, back-matching is provided. The HFVM is applicable to nearly all of the pixels, thus a dense disparity map is obtained in any case, and it is robust against

unknown scaling and rotation differences between the images to be matched. The proposed HFVM method requires 6.24 s for ME. In Figure (3.3) the resulting motion fields are illustrated.

3.2.4 FOLKI

The fourth algorithm is FOLKI [63], which is a local approach based on Lucas-Kanade (LK). It is a modified and convergent scheme for an accurate estimation including also large displacements. FOLKI performs fast correlation optimization over local windows by iterative registration. The used registration criterion is the displaced frame difference (DFD) criterion. Each component of optical flow is chosen in order to minimize this local criterion [63]. The minimization is achieved by using an iterative Gauss-Newton (GN) descent, as its convergence is known to be fast for small displacements [64]. Since larger displacements are possible as well, a multi-resolution scheme is applied in order to avoid local minima. An image pyramid is built by successively performing low-pass filtering and decimation. The GN iterations are initialized with a displacement equal to zero, which leads to first rough estimates. By descending the pyramid levels, those values are then refined one by one [64]. Such a coarse-to-fine multi-resolution scheme brings a dramatic improvement over an iterative one-resolution scheme and has also proven very efficient in many optical flow methods [63].

A key feature of FOLKI is its efficiency as it was designed as a massively parallel architecture. The three main types of computation (image interpolation, pixel-wise operations and separable convolution) can be accelerated on a graphics processing unit (GPU). FOLKI is a highly regular and parallel algorithm which is much more efficient than previous sparse LK techniques, and moreover, specially suited to GPU architectures. The computational costs depend on the window radius used for local registration and on the number of iterations at each image level [64]. The computed flow fields can be seen in Figure (3.4). On the GPU this algorithm takes 420.7 ms on average for one 2K frame.



(a) center → previous **(b)** center → next

Figure 3.4: Motion fields computed by FOLKI. On the left, the motion field, demonstrating the motion from frame 64 of sequence *Dance2* to the previous one. On the right, the forward motion, from center to the next frame is visualized.

3.3 Discussion

In this chapter four motion algorithms were discussed. Each of them can be applied to the single-frame-defect detection approach presented in this work. The evaluation of the proposed method in Chapter 4, will include two of them. Firstly, the TV-L1 optical flow is applied, as it uses a state of the art algorithm and produces a smooth motion field, shown in Figure (3.1). Due to the hardware requirements (GPU) of the TV-L1 optical flow, Pixel motion is provided as an alternative. It is able to utilize several CPU cores, thus an acceptable speed can be achieved. When comparing the results in Figure (3.2) with those of the GPU based optical flow, two differences are noticeable: The motion field which the TV algorithm produces is much smoother than that of Pixel motion, especially motion edges are detected precisely. Furthermore, the TV-L1 optical flow is able to deal with homogeneous regions, while the Pixel motion often deviates in such areas.

Although the resulting motion field of HFVM motion is similar to that of Pixel

motion, it is not further examined as its computational cost is too high and thus not applicable in the film restoration domain. Pixel motion has been used in DIAMANT for years, thus we concentrate on Pixel motion rather than on HFVM due to its interactivity in restoration tools. Finally, FOLKI has the best runtime, however, the estimated motion is much worse than that of the others.

As already mentioned, a detection algorithm requires a reliable ME, which is itself an ill-posed problem. It needs to know the motion at every position in an image, while this information is neither reliable in defected regions nor in occluded areas [40]. All of the described ME algorithms have difficulties in accurately matching such challenging parts. When one object occludes another, motion discontinuities are expected. Consequently, algorithms based on a smoothness constraint are likely to fail in areas of occluded edges [56]. As a result we have to deal with the fact, that optical flow algorithms only yield an approximate estimation of real motion. In general, the temporal and spatio-temporal methods, presented in Section (2.2.2) and (2.2.3) simply ignored the fact that ME can fail and that occlusions can occur. Thus, there is a need to improve the robustness of applications dependent on ME, especially those in the restoration domain [55]. This can be achieved by detecting regions of unreliable ME and by treating such regions as special cases [55]. Our approach is based on this idea. Our main objective is to consider motion failures in the detection approach, which is further examined in the following chapter.

Chapter 4

Dirt and Dust Detection Approach

4.1 Introduction

In this master's thesis, we propose a novel approach based on the **r**eduction of **FA** by analyzing the **M**otion estimation followed by an **E**xtraction of **D**amping functions (rFAMED) in order to provide reliable dust masks.

The quality of an obtained dust mask can usually be judged by two quantities, the good or correct detections (CD) and the false detections or false alarms (FA). Desirably, the resulting dust mask has a high number of good detections, since every missed defect won't be restored and at the same time a low number of false detections, since FA endanger artifacts. Basically, related approaches regularize these two terms by a chosen threshold. Mostly, the right balance cannot be found, since dependent on the characteristic of the film material, either the FA are over-balanced or an acceptable CDR cannot be afforded. More recent work concentrates on post processing where falsely detected defects are identified and removed from the dust mask in a following step. Due to thresholding of the dust mask, valuable information of each pixel is lost. A better way is proposed by Ren and Vlachos in [25], where they extract a confidence signal. This confidence measure is a probability measure for each pixel indicating the likelihood of being defected. By keeping this information, FA can be controlled in a post processing step.

The presented approach was chosen systematically, since we first figured out, why FA occur. Possible reasons are noise, motion estimation (ME) and accordingly warping errors, and occlusion/uncovering areas. Related work regarding analysis and detection of pathological motion provide quite good results. As a result, we have developed a novel dirt and dust detection technique as a three-step approach:

1. Primary dust response: In contrast to most related work we first infer response values for each pixel instead of making an early binary decision. For the primary measure, we use a sROD based detection method, since simplified rank-ordered difference (sROD) has been excessively used in recent work and has outperformed other techniques.

2. Noise suppression: The so-called co-support operator is presented. It "supports" pixels that are in a blotchy neighborhood by increasing their likelihood to belong to a dust spot while reducing response values of pixels that are in a clean neighborhood.

3. Damping functions: Several novel measures are applied in order to find areas of pathological motion (PM), since motion cannot be reliably estimated in such regions. The functions are mainly extracted from motion field analysis. Responses which reside in these areas are preemptively damped down, in order to avoid further disturbing artifacts caused by the replacement of true image data during the removal process.

In the following subsections each step of the dirt and dust detection algorithm is illustrated and explained in detail. At the end of this chapter the final combined approach is presented, consisting of empirically selected methods which appeared to produce the best results.

4.2 Primary dust response measure

The first step of our novel approach is to obtain the primary dust response measure. Similar to the confidence measure of Ren and Vlachos [12], the detector response R is indicating the likeliness of being defected for each pixel. Basically, it is a difference of two corresponding intensity values.

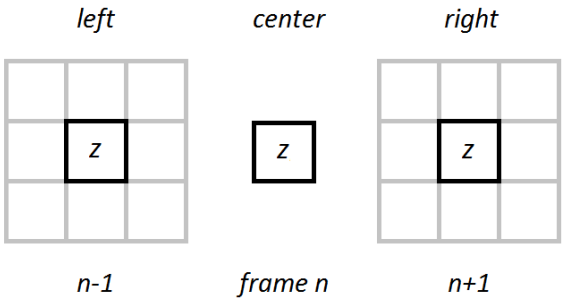


Figure 4.1: Spatial and temporal neighborhood of z_n. The spatial neighborhood is restricted to the actual frame, the temporal neighborhood only includes pixels from neighboring frames, that have exactly the same position as the actual observed pixel. The spatio-temporal neighborhood involves both, the temporal neighbors and the appropriate neighborhood of the temporal neighbors.

Considering three consecutive frames, for each pixel $z_n = (x, y)^T$ of the middle frame n, a response measure $r(z_n)$ is inferred by including the temporal and spatial neighborhood, as shown in Figure (4.1). This measure allows to keep important information of each pixel, and later it is used to infer the defect-probability. Due to the requirement of a low computational cost, an efficient method is needed to compute the primary dust response measure.

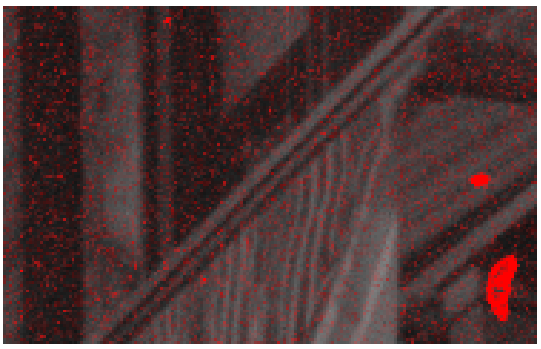
Pixel based methods, such as SDIa, SDIp and sROD, are very fast, have good detection results and can easily be adapted in order to provide probabilities instead of binary values. For the SDIa based method [20] for example, the response value $r(z)$ for the current pixel in frame n is obtained by using combined differences as defined in the following:

$$r_{n,SDIa}(z) := min\{|I_n(z) - C_{n-}(z)|, |I_n(z) - C_{n+}(z)|\} \tag{4.1}$$

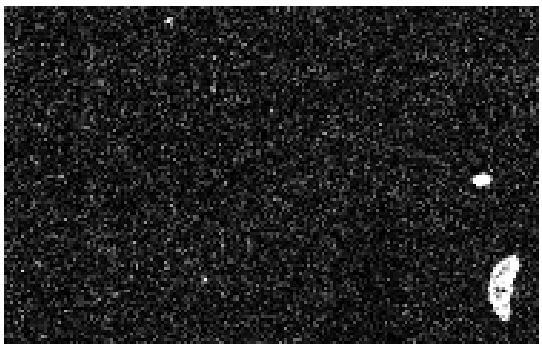
where C_{n-} and C_{n+} are the motion compensated previous and next frames respectively. The result of the SDIa based method is illustrated in Figure (4.2).

In order to gain probability values and also for visualization purposes, seen in Figure (4.2 - 4.5), a r_{max} is defined. All response values are clipped to a range of

(a) Original image



(b) Response encoded in red channel



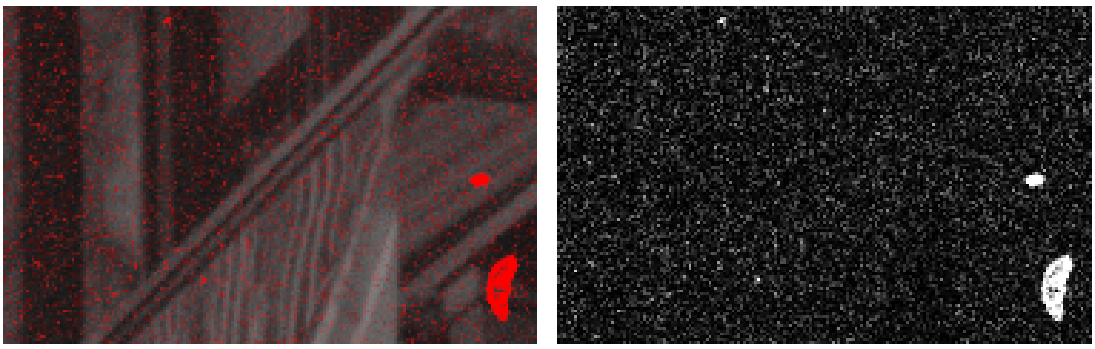
(c) Response encoded as intensities

Figure 4.2: SDIa based response measure, indicating the probability of each pixel of being defected. (a) a zoomed part of the original frame taken from sequence *JazzMan*, (b) probabilities are distributed over the red channel and overlaid with the gray scaled center image, (c) probabilities are visualized as an 8-bit gray-scaled image, distributed over the intensity channel.

$[0, r_{max}]$ and are linearly transformed into a probability measure of $[0, 1]$.

Due to the sign constraint, the SDIp detector [20] produces lower FAR than the SDIa. A penalty term for different signs can be introduced in our formulation as well. The basic idea is, to dampen down the response value of the current pixel if the forward and backward differences do not have the same sign. Different signs indicate the presence of noise or motion estimation (ME) errors. Thus, Equation (4.1) is extended to

$$r_{n,SDIp}(z) := min\{|D_-|, |D_+|\} \cdot ((1 - \lambda)\ sign(D_-, D_+)), \tag{4.2}$$



(a) Response encoded in red channel **(b)** Response encoded as intensities

Figure 4.3: SDIp based response measure: Additionally, the sign consensus of SDIp is integrated in the response measure.

where $\lambda \in [0, 1]$ defines the damping factor, $D_- = I_n(z) - C_{n-}(z)$ is the backward difference, $D_+ = I_n(z) - C_{n+}(z)$ is the forward difference and *sign* is a function that returns 1 if the signs of the differences are equal or 0 otherwise. The resulting response is illustrated in Figure (4.3).

In order to take the spatial neighborhood of the previous and the succeeding frame into account, the sROD detector [6] can be adapted as well. Let

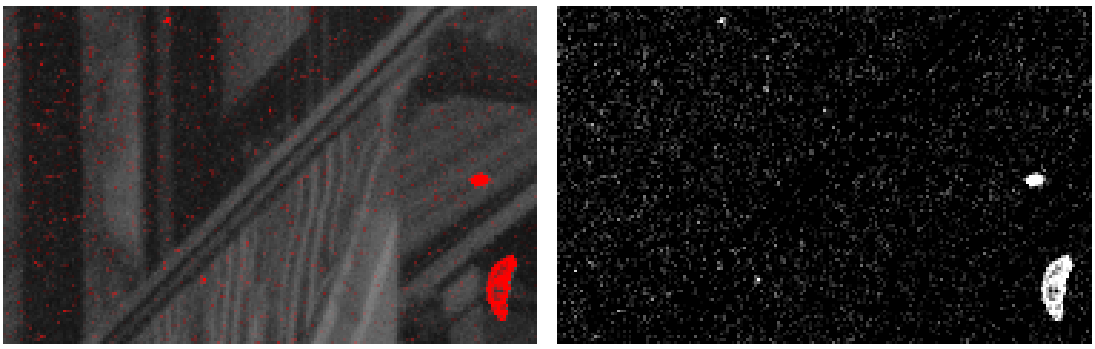
$$N_{LR} \subseteq (N_8(z_{n-1}) \cup N_8(z_{n+1})),$$

where $N_8(z_{n-1})$ and $N_8(z_{n+1})$ is the 8-neighborhood of z_{n-1} and z_{n+1} respectively. Furthermore, $v_i \in N_{LR}$ and s is the min difference of the center element and each of its neighbors. Thus, by using

$$t := r_{n,SDIa} \text{ and} \tag{4.3a}$$

$$s := min\{|v_i - z_n| \mid \forall v_i \in N_{LR}\}, \tag{4.3b}$$

we obtain the result for the sROD based method by



(a) Response encoded in red channel **(b)** Response encoded as intensities

Figure 4.4: sROD_2 based response measure: Two vertical neighbors are additionally considered in each temporal neighboring frame.

$$R_{sROD} := \min\{t, s\}. \tag{4.4}$$

The result for the sROD_2, which includes the two vertical neighbors (classical variant described in [6]), is illustrated in Figure (4.4).

Any possible configuration of neighbors can be considered when using this formulation. The more neighboring pixels are included, the stricter the response measure turns out to be. To regularize the strictness of the response measure we introduce an additional term λ. A general sROD based response measure R_{gsROD} is then defined as follows

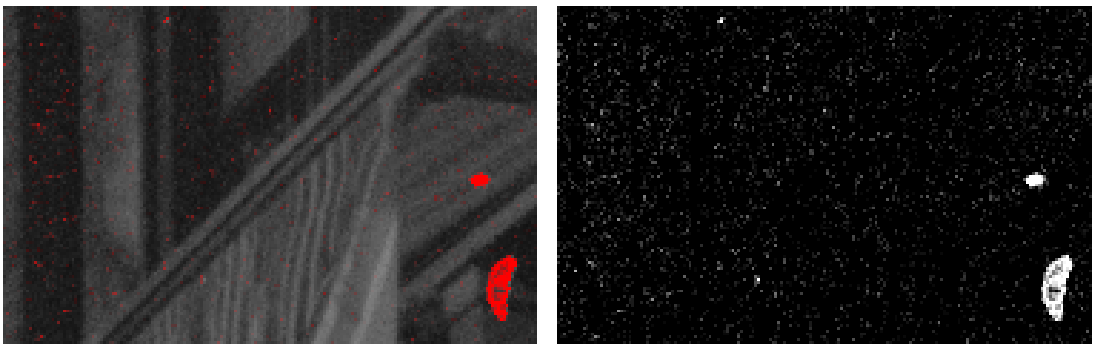
$$R_{gsROD} := \min\{t, \lambda \cdot s\}, \tag{4.5}$$

where $\lambda \geq 1$. In Figure (4.5) the R_{gsROD} examining first order neighborhood (4-neighborhood) of each reference pixel, z_{n-1} and z_{n+1} is illustrated.

As a consequence,

$$\begin{aligned} \lambda = 1 \quad &\rightsquigarrow \quad R_{gsROD} = R_{sROD}, \\ N_{LR} = \emptyset \quad &\rightsquigarrow \quad R_{gsROD} = R_{SDIa} \end{aligned}$$

and



(a) Response encoded in red channel **(b)** Response encoded as intensities

Figure 4.5: sROD_4 based response measure by using 4-neighborhood: Two vertical and horizontal neighbors are additionally considered in each temporal neighboring frame.

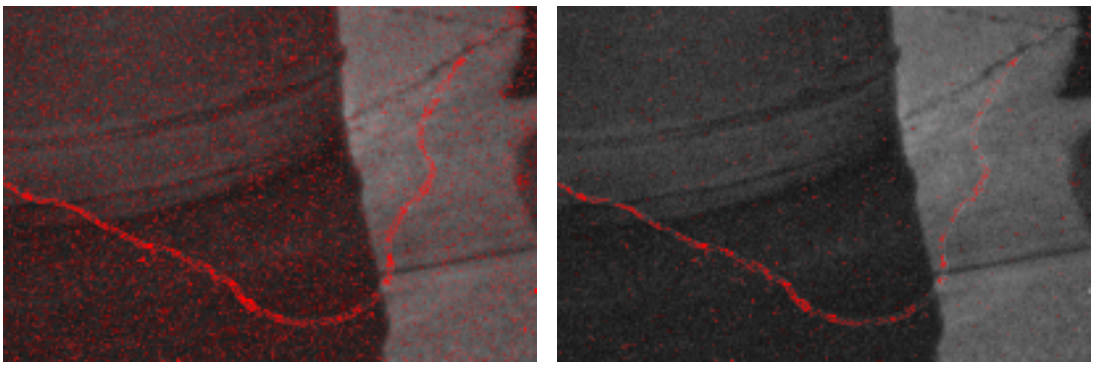
$$R_{sROD} \leq R_{gsROD} \leq R_{SDIa} \quad \text{if} \quad N_{LR} \text{ is fixed,}$$
$$R'_{gsROD} \geq R_{gsROD} \quad \text{if} \quad N'_{LR} \subseteq N_{LR}.$$

Thus, the general sROD, taking all 8 neighbors into account, is most robust against single outliers. It means, this method has the least false alarms (FA) due to noise. However, as shown in Figure (4.6), when considering an 8-neighborhood, the detection of fine structured dust spots turns out to be insufficient. Although, noise is significantly reduced, fine structures are not detected in its whole extend either. The best balance between reduction of noise and preserving fine structured dust spots has to be found. Therefore, the sROD_2 or sROD_4 based response measures, shown in Figures (4.4) and (4.5), seem to be the better choice.

4.3 Noise suppression

As already shown in the previous section, the presence of noise is an substantial problem in film restoration applications. A common consequence of the single-frame-defect detection is a high FAR. Pixel based detectors are not able to distinguish between noise and small dust spots or low contrast blotches, which challenges the detection process.

In Section (4.2) it was already shown, that FA due to noise can be significantly



(a) SDIa based response measure **(b)** $sROD_8$ based response measure

Figure 4.6: Comparison between SDIa and $sROD_8$ based response measure, showing the improvement regarding noise suppression due to considering more spatial information.

reduced by including a certain spatial neighborhood. However, when considering a high number of neighbors, response values of fine structured dirt and dust are damped down radically as well. In this section an efficient technique is proposed in order to counteract this effect.

4.3.1 Co-support operator

The co-support operator was motivated by the characteristic of blotches, that dust spots form coherent regions and that the occurrence of single dust spots is rather rare. Similarly to the post processing methods proposed in [6], we need to keep the FAR due to noise low, dust spots have to be completed if single pixels within a dust spot have low responses and an overall dilation is advantageous for the border cases arranged at dust spot edges. All these steps are combined in one with the co-support operator, which further shares some similarity with the histogram of template (HOT) features, proposed in the work of Tang and Goto [65]. In their work they propose a novel feature for human detection in still images, named HOT. In a 3×3 pixel region various templates are given in order to define the positional relationship of three pixels, the center pixel and two of its neighbors. The HOT feature verifies if pixels meet one of the predefined templates by using intensity

and gradient information.

We follow a similar approach by using only intensity information. The co-support operator adapts pixels by checking for similarity to its neighbors, considering also a 3×3 pixel region. Basically, the idea of this operator is the following: The response value of a pixel and consequently its probability to be part of a defect should be increased if the pixel is a neighbor of dust spot pixels that are supposed to have a high response. A dust spot in the center frame which does not have overall significant response values, is triggered to be "filled up". Vice versa, if pixels do not have neighbors with high responses, they are likely to be outliers. In such a case, the operator will decrease the response. The resulting effect is, that noisy, small-sized outliers are eliminated if their response is not significant.

General co-support operator C_k

Let v_c be a value, representing e.g. the response, intensity, homogeneity, etc., of a pixel at position $z = (x, y)^T$ and $v_n \in N(v_c)$ is the value of each of its n neighbors. Furthermore, $f(\cdot, \cdot)$ computes a similarity measure and function g is used to adapt the value of the center pixel due to the observed neighborhood. Mainly, g will be set to min_k or max_k, which represents the k-highest or k-smallest value, respectively.

Then, the general co-support operator $c_k(z)$ is defined as

$$c_k(z) = g\{f(v_c, v_n)|v_n \in N(v_c)\}, \tag{4.6}$$

Additionally, this operator can be extended in such a manner as to consider the direction of v_c to v_n, as it is done in the work of Tang and Goto [65]. In this way, the comparison of center and each neighbor can be weighted individually.

Co-support operator for dirt and dust detection

The co-support operator c_k is applied to the response measure in order to improve the response measure in relation to noise and incomplete dust spots. The challenge is to choose an adequate operator f. Choosing *max* or *average* as similarity operator f, the result will show a growing effect. Vice versa, *min* would result in an erosion effect.

Thus, we define a weighted combination of *min* and *average*:

$$f(r_c, r_n) = \alpha_1 \cdot (min(r_c, r_n)) + \alpha_2 \cdot (\phi(r_c, r_n)),\ \forall r_n \in N_8(r_c) \tag{4.7}$$

where ϕ is the average function: $\phi = \frac{r_c + r_n}{2}$. Finally, the updated value is obtained by using $c_k(z) = max_k\{f(r_c, r_n)\}$.

As already explained, the response $r(z)$ is obtained by using combined differences such as e.g. the SDIa based method, see Equation (4.1). We clip the response values to a range $[0, t \cdot \tau]$ and subtract τ in order to obtain response values in a range of $[-\tau, (t-1) \cdot \tau]$ by using

$$r'_c(z) = min(r(z), t \cdot \tau) - \tau, \tag{4.8}$$

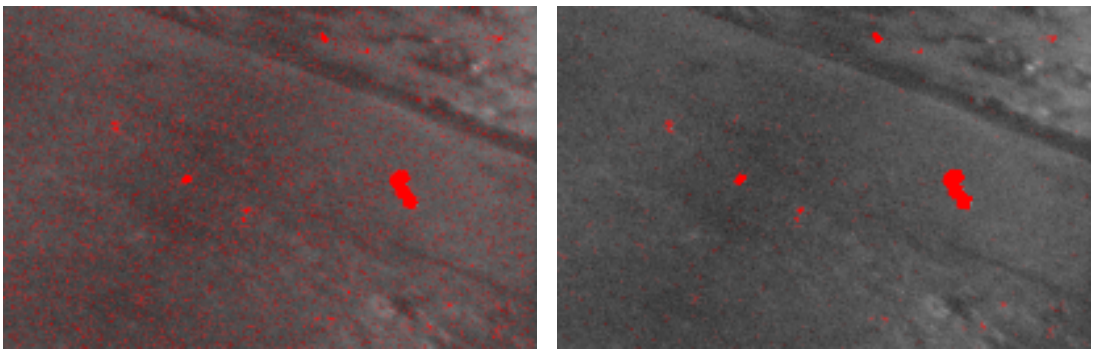
where τ is the dust threshold. Therefore, negative values indicate pixels which are rather no dust spots, positive values correlate with pixels that are likely to belong to a dust spot, and values around 0 are border cases. It appears to work best to set t to a value of 2 in order to limit the response values to $2 \cdot \tau$. The reason for that will be clear later on, when we introduce the damping functions.

Ideally, the similarity function $f(r_c, r_n)$ should correspond with function $min(r_c, r_n)$, if r_c is rather low, i.e. $r_c < 0$, and it should be $\approx \phi(r_c, r_n)$ if r_c is significant, i.e. $r_c > 0$, as at this point the probability of pixels being a dust spot increases.

Let $d = |\min(\max(r'_c, -\tau), 0)|$ be a function to minimize the 'growing effect' at boundaries, which will be used to compute the weights for the co-support operator. Consequently, function f which has to be computed for every neighbor $r_n \in N_8(r_c)$ is defined as

$$f(r_c, r_n) = \frac{d}{\tau} \cdot min(r_c, r_n) + (1 - \frac{d}{\tau}) \cdot \phi(r_c, r_n). \tag{4.9}$$

Note that the co-support operator mainly damps responses of outliers and slightly increase response values of pixels, which have a neighborhood of high responses. Due to the *average* function, the increasing of responses will always be limited to the average of center and highest neighbor response, which may be too weak. We



(a) Response based on SDIa **(b)** Result of co-support operator

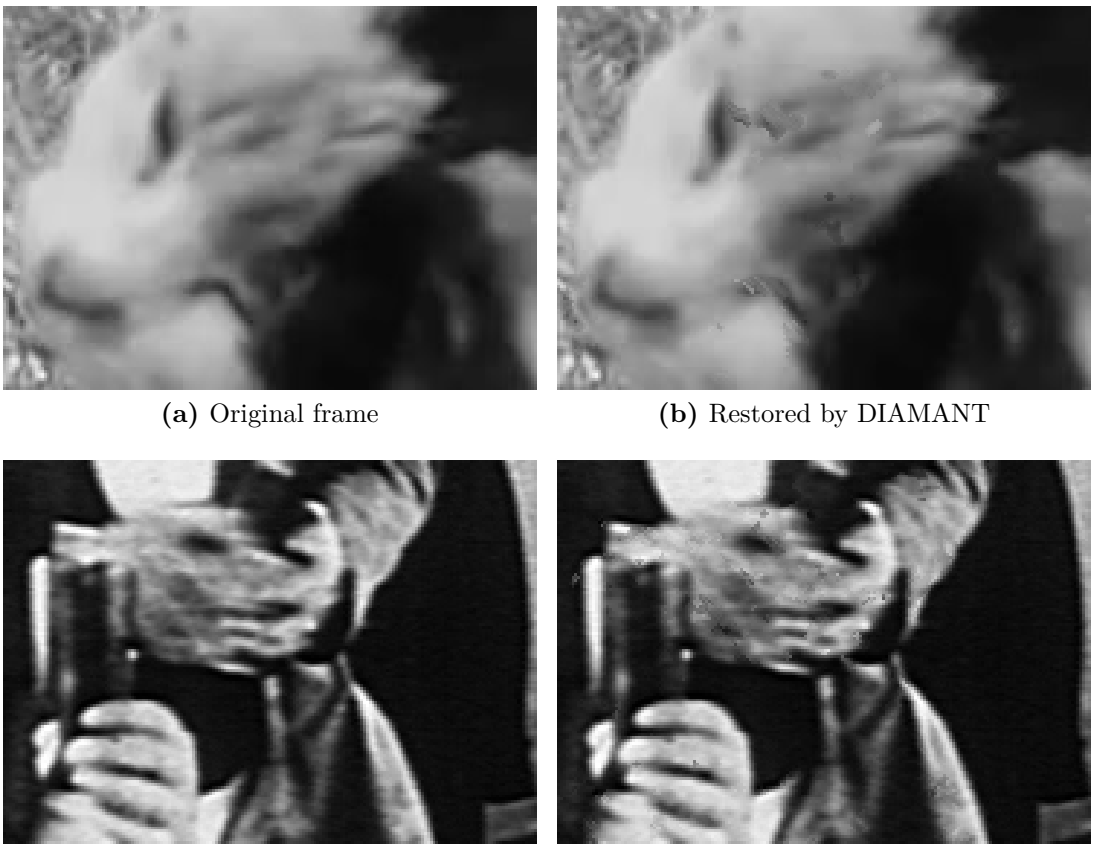
Figure 4.7: Resulting effect of the co-support operator. Comparison of SDIa based response, in (a) without and in (b) with co-support operator applied.

have found that a non convex weighting function, where the *average* function is constantly assigned a higher weight than the *min* function, will counteract this issue.

In Figure (4.7) the result of the co-support operator is demonstrated. Single outliers in the response measure are effectively reduced, while the response values of small, higher contrasted parts are not decreased. Additionally, as desired, the response values at the border of potential dust spots are effectively increased.

4.4 Damping functions

Most false alarms (FA) occur due to pathological motion (PM), close to moving edges or simply as a result of ME errors. Consequently, areas, that are likely to result in false detections, have to be found and treated separately. By a comprehensive analysis of the motion field, provided by the used ME algorithm, and the appropriate warped neighboring frames, we derive measures, that indicate precarious areas. Precarious areas are occlusion or uncovering areas, moving edges, areas of complex motion, etc. The obtained measures are transformed to a probability distribution for each pixel and further utilized to damp down the response values in the precarious regions. Finally, the response measure is thresholded in order to



(a) Original frame **(b)** Restored by DIAMANT

(c) Original frame **(d)** Restored by DIAMANT

Figure 4.8: Samples for undesired restoration artifacts: Due to ME errors, particularly in areas of complex motion, artifacts may be introduced to the restored result.

obtain a binary detection mask. In the removal process, the detected image regions are replaced by using information of the motion compensated neighbors. If defects are detected due to motion estimation (ME) errors, image data is replaced by values from image parts that are not reliably matched. The obvious effect is the generation of artifacts, which is illustrated in Figure (4.8). For those samples, the detection as well as the removal of defects was done by the DIAMANT film restoration algorithm.

We define X measures in order to assign an appropriate probability to all parts of the current image. Each of the measures is presented in detail in the following subsections. For each measure $x \in X$, 'dangerous' regions (i.e. PM, unreliably estimated motion, moving edges, etc.) are tagged with high probability values $p \in [0, 1]$. The final probability distribution is then obtained by combining all x measures using

$$p_{final} = \prod p_x, \tag{4.10}$$

where p_x is the probability distribution of each single measure x. In the last step we multiply the inverse damping probability p_{final} by the obtained response. Ideally, FA due to ME errors and due to PM are now below the threshold τ and consequently, are not detected. The damping function approach obviously does not improve detection accuracy, but it represents an effective mechanism for controlling FA.

As mentioned earlier, the detection of dirt and dust is particularly challenging in areas of complex motion, since ME algorithms are likely to fail in such areas. The idea of our approach is to analyze the motion field in both directions, from the center to both neighboring frames. In this way we find regions in the current image which were not estimated reliably.

In the following we list the used measures, which can be combined in the end and consequently indicate parts of the frame, where blotch candidates are likely to be detected due to ME errors. In such areas we rather damp down the response values than clear the detection mask, as it should be still possible to detect high contrasted blotches. Some of the following measures are exclusively inferred from the motion field, while others are computed by including also the center and warped neighbors. The following damping functions can be computed either as a combination of both directions (backward and forward) or by choosing one direction. If only one direction is considered, the function might be less reliable and partial, but the computational cost is lower. In order to depart from single-pixel-decisions we apply a box filter of varying size to all damping measures.

As already mentioned in the previous chapter, we mainly use TV-L1 optical flow and Pixel motion for ME. Thus, each of the visualizations of the damping function measures include both algorithms.

4.4.1 Motion velocity

We assume that an object movement from one frame to the next one is finite, dependent on the image width. Excessively large motion vectors are commonly measured if the current position is arranged in an occluded region and the referring position cannot be found or if the motion is very complex. Therefore, motion vectors exceeding a predefined v_{max} are likely to be wrongly estimated.

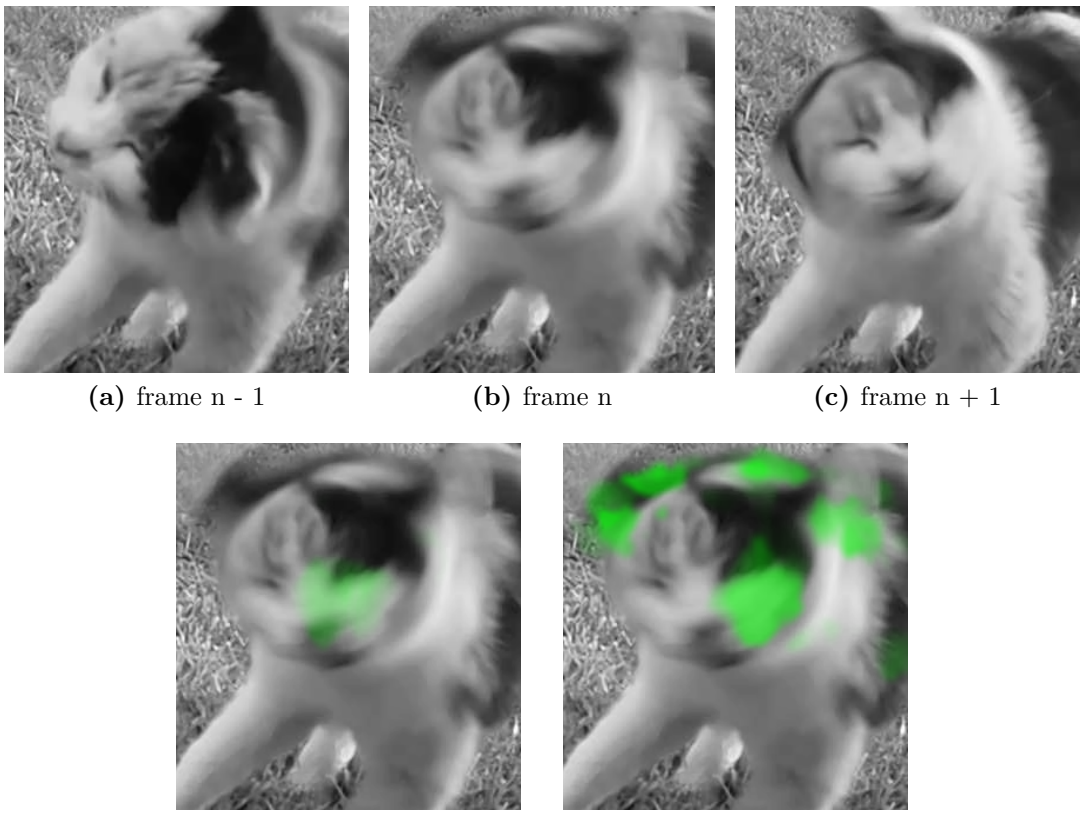
Let $v(z)$ be the magnitude or also denoted as velocity of the motion vector at spatial position $z = (x, y)^T$, where the range $[0, \epsilon]$ represents proper velocities of a motion vector, and let v_{max} be the maximum velocity an object is supposed to move within two consecutive frames.

The probability of a pixel being unreliably estimated due to a high magnitude of its motion vector, $p_{velocity}(z) \in [0, 1]$, is then obtained by

$$p_{velocity}(z) = \omega_{velocity} \cdot \min\{\frac{\max\{v(z) - \epsilon, 0\}}{v_{max} - \epsilon}, 1\} \tag{4.11}$$

The damping can still be regularized by $\omega_{Velocity} \in [0, 1]$. For the Pixel motion we use a slightly bigger ϵ as the Pixel motion is less smooth and therefore has more outliers. The resulting probability density function is illustrated in Figure (4.9), where the probability values $p_{velocity}$ are encoded in the green channel and are overlaid with the center image. For both ME methods, the velocity measure shows similar results. As shown in the image sequence, the cat shakes its head very fast, so the right parts of the image are affected.

This measure has to be applied considerably, since all defects that are placed on the moving object might be damped down as well, although the reference pixel are available in both directions. Another issue of this measure is, that fore- and background are not distinguished. If fast motion occur due to camera motion, the whole image might take effect. Thus, assigning a low $\omega_{Velocity}$ might be advantageous instead of a higher one, as real blotches would be damped down radically as well.



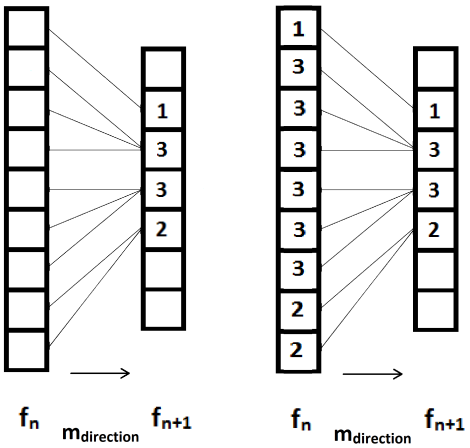
(a) frame n - 1 **(b)** frame n **(c)** frame n + 1

(d) Result using TV-L1 optical flow **(e)** Result using Pixel motion

Figure 4.9: Probability measure inferred from the velocity of motion vectors, illustrated for image sequence *Cat*. (a) - (c) shows the three succeeding image frames, (d) shows the magnitude measure when using TV-L1 optical flow and in (e) the measure is shown when Pixel motion is used for the ME.

4.4.2 Density of motion vector fields

The density of a motion vector field offers valuable information regarding occlusion and uncovering areas. This finding was already utilized in the literature in a similar way, e.g. in the work of Bartels et al. [66]. If several motion vectors point to the same position in one particular region, high density is observed. High density indicates an occlusion area. As the actual reference pixels are occluded, all motion vectors point to the boundaries of the occluded area. Vice versa, regions have



(a) Step 1 **(b)** Step 2

Figure 4.10: Computation of density measure, by applying 2 steps: In the first step, the motion vectors are followed and the number of references are accumulated for every pixel. In the second step, the number of references are inserted back into the appropriate position of the center frame.

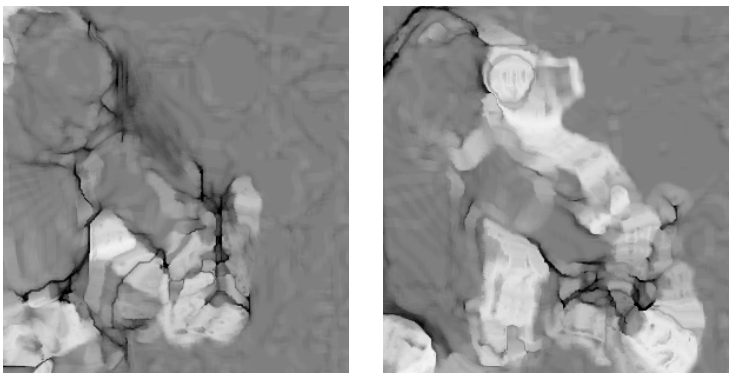
low density if no or few motion vectors are pointing to it. Low density indicates areas of uncovering. The computation of the density image is illustrated in Figure (4.10). In a first step, an 'accumulator image', denoted as num_{ref}, is built, where the number of references is incremented at the neighbor's target positions by using

$$\forall z: \; num_{ref}(z + \vec{m}_z) = num_{ref}(z + \vec{m}_z) + 1,$$

where $z = (x, y)^T$ is the pixel position in the center frame and $\vec{m}_z$ its motion vector. The indexes of the accumulator image correlate with those of the neighboring frame. Then num_{ref} is slightly blurred by applying a box filter. In the second step, the density image is constructed by taking the corresponding values from the accumulator image, as explained in Figure (4.10). The indexes of the density image correlate with those of the center frame.

The density image $m_d(z)$ is obtained by using the following equation:

$$m_d(z) = f_b(num_{ref}(z + \vec{m}_z)), \tag{4.12}$$



(a) backward direction **(b)** forward direction

Figure 4.11: Logarithmic visualization of expanded and compressed parts, indicating the density of the motion field. For this visualization the motion was estimated by using TV-L1 optical flow.

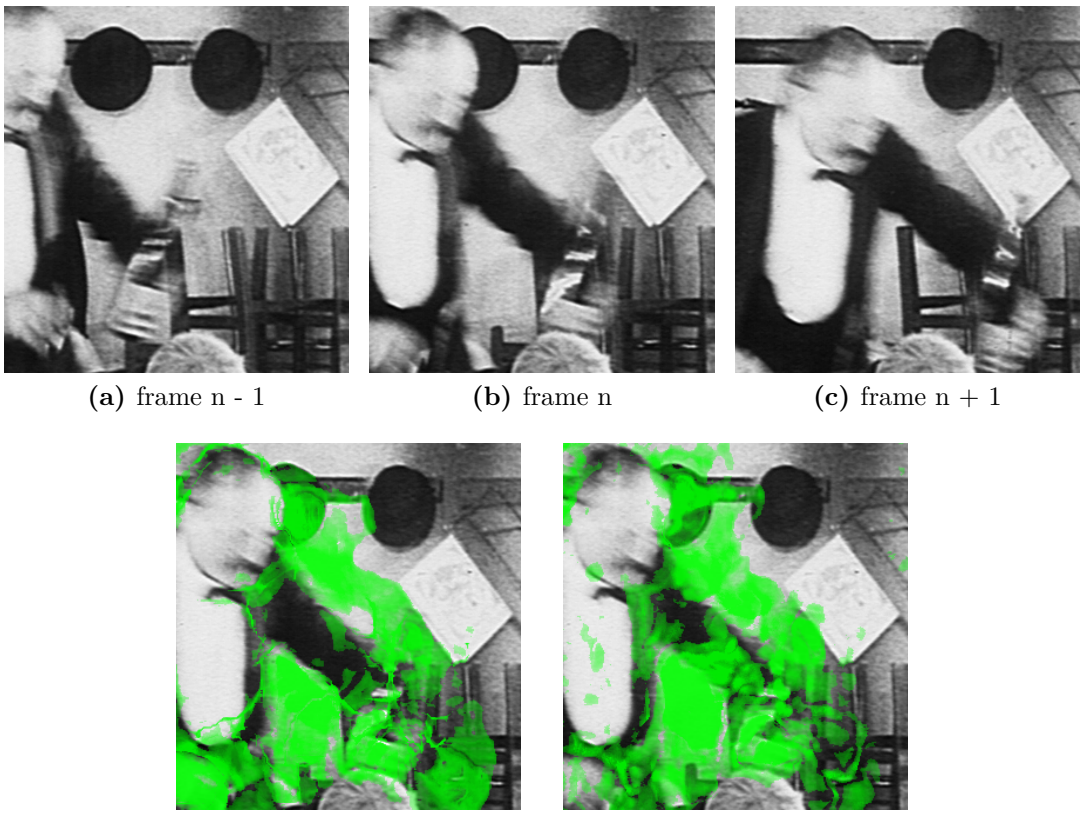
where f_b is a box filter, $num_{ref}(z)$ indicates the number of references at position z in the neighboring frame f_{n+1}.

Finally, the density image m_d is clipped to the range $[\frac{1}{m_{d_{max}}}, m_{d_{max}}]$, where $m_{d_{max}}$ defines the max references, which correlates with a probability of 1. For illustration purposes a logarithmic scaling is applied, see Figure (4.11), where all bright values correlate with expansion and all dark values with compression of the presented motion. All values referenced only once are assumed to be reliably estimated. Other parts denote areas of occlusion or uncovering, since no unique reference could have been found for these positions. The probability of a pixel lying in an occlusion/uncovering area due to high/low density values is finally computed by

$$p_{density}(z) = \omega_{density} \cdot \frac{|\log(m_d(z))|}{\log(m_{d_{max}})}, \tag{4.13}$$

where $\omega_{density}$ is the damping weight for the density function.

In Figure (4.12) the probability distribution $p_{density}$ is visualized, encoded in the green channel. The density, estimated by the TV-L1 optical flow, see Figure (4.12d), is accurately estimated due to the smoothness of the algorithm. It is slightly better than the result obtained when using Pixel motion, as shown in



(a) frame n - 1 **(b)** frame n **(c)** frame n + 1

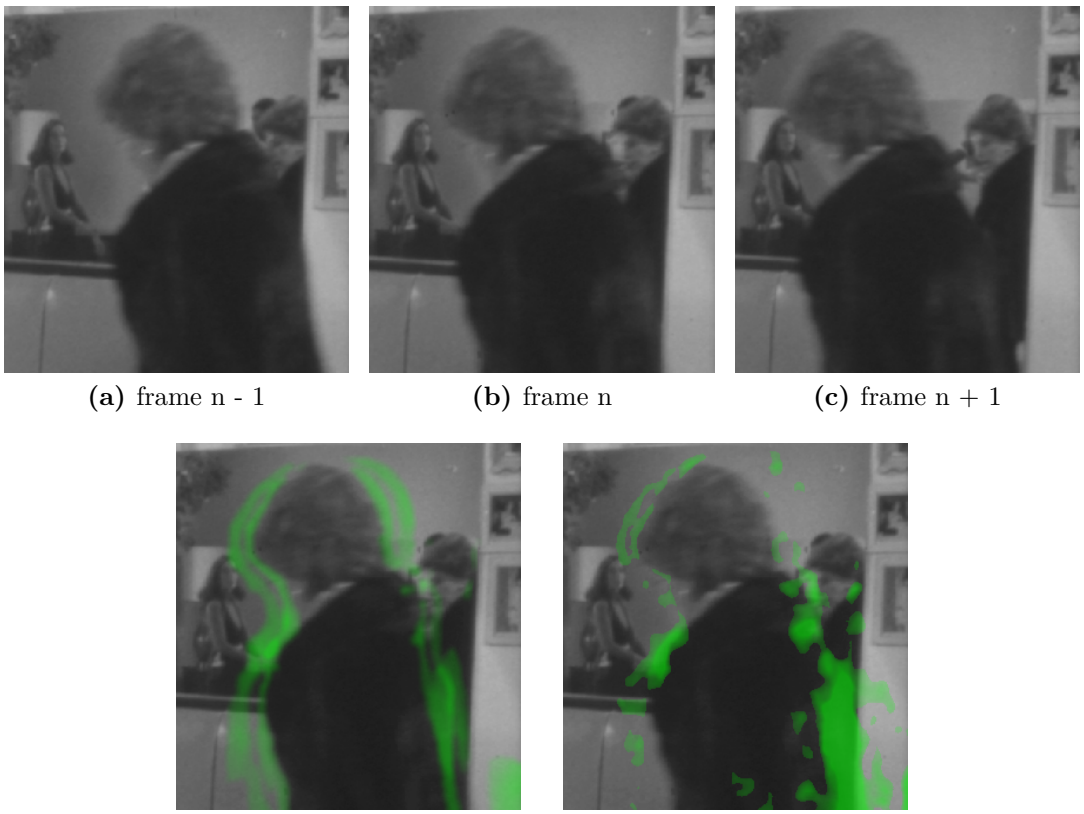
(d) Result using TV-L1 optical flow **(e)** Result using Pixel motion

Figure 4.12: Probability measure inferred from density of motion field, illustrated for image sequence *Stadt ohne Juden*. In (d) and (e) it is shown, that the occlusion and uncovering areas are correctly affected for both ME algorithms.

Figure (4.12e). The calculated density using Pixel motion may be badly affected if illumination changes are present, since the Pixel motion cannot deal with illumination changes and produces more ME errors.

4.4.3 Back-matching of motion vectors

For the back-matching measure the motion vectors have to be available in both, the forward ($\vec{m}_f$) and the backward ($\vec{m}_b$) directions. A reliably estimated motion vector is assumed to be equally measured in both directions, which means that



(a) frame n - 1 **(b)** frame n **(c)** frame n + 1

(d) Result using TV-L1 optical flow **(e)** Result using Pixel motion

Figure 4.13: Probability measure obtained by disparities of back-matching measure, illustrated for image sequence *Surprise*. The back-matching measure affects occlusion and uncovering areas and, in particular, areas of ME errors.

the following assumption must hold

$$\forall z_c : z_n = z_c + \vec{m}_f(z_c) \text{ AND } z_c = z_n + \vec{m}_b(z_n), \tag{4.14}$$

where z_c is the actual position in the center frame and z_n is the corresponding (indicated by the motion vector), matched position in the neighboring frame. If this assumption is violated, the motion has not been accurately estimated, which may happen due to complex motion, occlusion, uncovering, etc.

Let u and v be the disparities in the horizontal and vertical direction, respec-

tively, and let ϵ be tolerance value, the probability of a pixel being unreliably estimated, measured by back-matching disparities, can be determined as follows:

$$p_{bm}(z) = \omega_{bm} \cdot \min\{\frac{\max\{f_b(\max\{u(z), v(z)\}) - \epsilon, 0\}}{m_{bm_{max}} - \epsilon}, 1\}, \tag{4.15}$$

where f_b is a box filter, $m_{bm_{max}}$ is the max defined disparity which will be transformed into a probability of 1 and ω_{bm} is the damping weight. The probability distribution is linearly scaled in the range of $[\epsilon, m_{bm_{max}}]$.

Back-matching is a very effective feature to detect ME errors. Its drawback is that the motion has to be calculated in both directions, which has a negative impact on the computational cost. However, if an automatic restoration mode is provided, motion does not have to be calculated additionally. It is rather an issue of book-keeping, since the motion vector fields have to be saved over a longer period of time. In Figure (4.13) the probability distribution p_{bm} is illustrated.

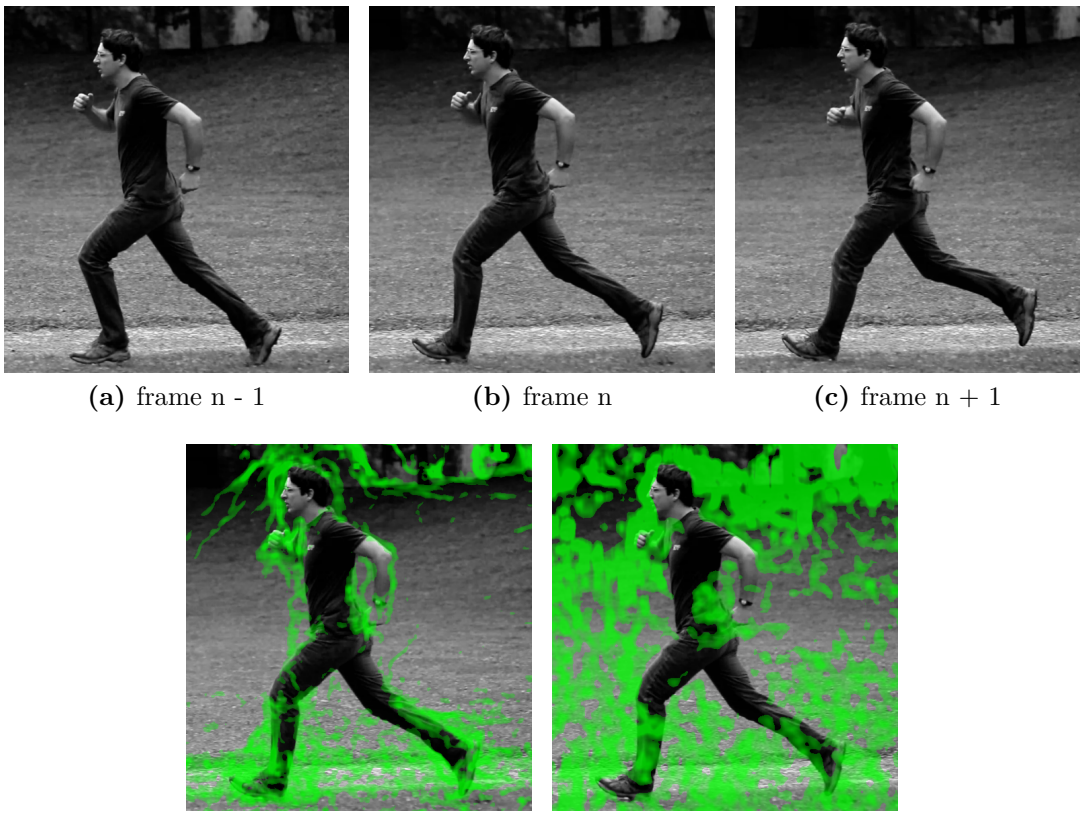
The back-matching function mainly indicates ME errors as well as occluded and uncovered parts of the image, since in such areas either the forward or the backward target positions of the motion vector is unavailable.

4.4.4 Divergence of motion vector field

The measure of divergence was inspired by Corrigan et al. [44]. According to Corrigan et al. PM and ME failures could be diagnosed by concentrating on the motion field itself, since PM often violates the local smoothness assumption of a motion field. In their work the vector field divergence is used as it is a suitable measure of smoothness within a vector field. High divergence values correspond to regions of the field which are not smooth and so may be associated with PM.

The divergence measure is computed by including both components of the motion vectors, the u- and v-translation vectors. Basically, the divergence of optical flow is the sum of the derivatives in two perpendicular directions, u- and v-direction, and indicates motion boundaries. Therefore, the higher the difference between the motion of each side, the higher is the divergence value. A box filter is applied to this measure in order to slightly extend the motion boundaries.

When viewing the outcome, see Figure (4.14), a noticeable thing is that when Pixel motion is used, the result is much less accurate. This can be explained by



(a) frame n - 1 **(b)** frame n **(c)** frame n + 1

(d) Result using TV-L1 optical flow **(e)** Result using Pixel motion

Figure 4.14: Probability measure obtained by divergence of motion field, illustrated for image sequence *Hannes runs*. The divergence measure mainly affects motion boundaries. For the Pixel motion, this measure is less applicable.

the fact that the motion field produced by Pixel motion is more erratic.

The divergence of a vector field is a scalar field and in the two dimensional case it is defined as:

$$div(u, v) = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}. \tag{4.16}$$

Finally, the probability of each pixel, to be arranged in an area of ME errors due to high divergence values, is defined as follows:

$$p_{div}(z) = \omega_{div} \cdot div(u, v), \tag{4.17}$$

where ω_{div} is the damping weight for the divergence measure. In Figure (4.14) the probability values are visualized, encoded in the green channel.

4.4.5 Difference of motion-compensated neighbors

If the motion is accurately estimated, the motion-compensated neighboring images and the center frame are supposed to be equal. Since the center frame possibly includes dirt and dust, while the neighbors are assumed to be clean regarding defects, the DFD of the two warped neighboring frames indicate ME errors.

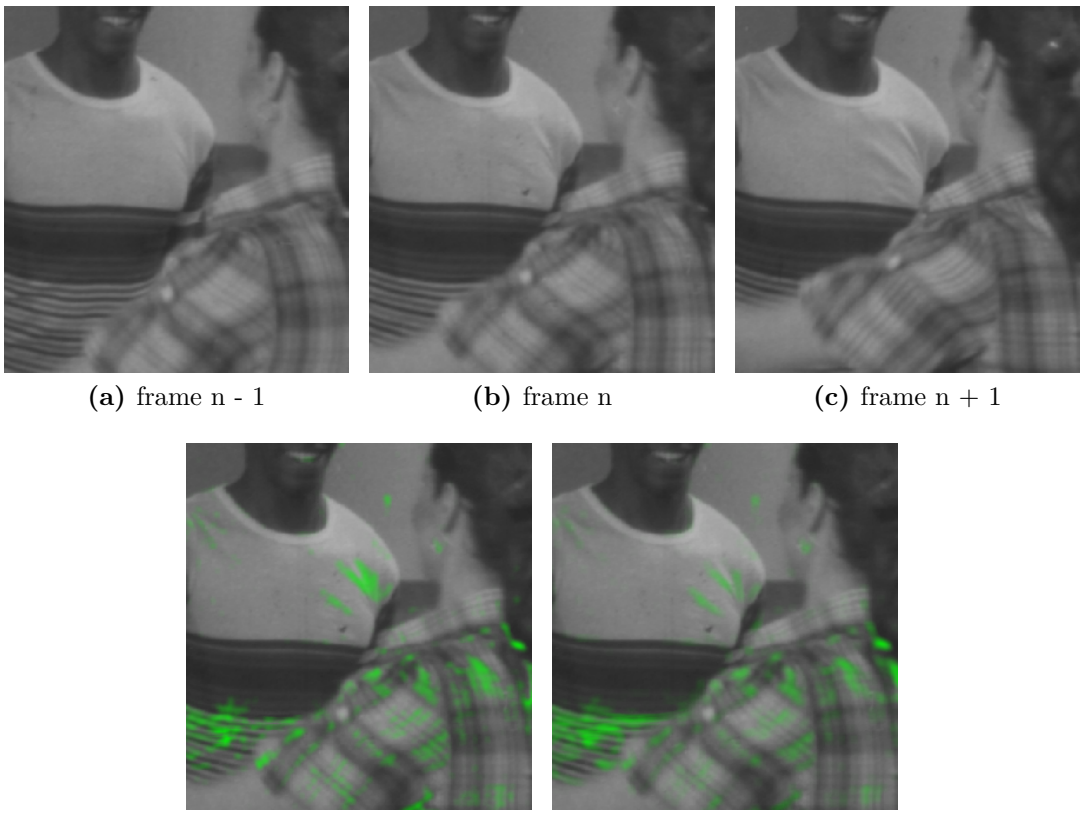
This measure is similar to the constraint used for the SDIjrs method, described in Section (2.2.2), where this additional constraint of similarity of warped neighbors already effectively reduced FA. The corresponding probability of each pixel is obtained by

$$p_{warpedDFD}(z) = \omega_{warpedDFD} \cdot \min\{\frac{\max\{|C_{n-}(z) - C_{n+}(z)| - \epsilon, 0\}}{m_{w_{max}} - \epsilon}, 1\}, \tag{4.18}$$

where $m_{w_{max}}$ is the maximum difference which corresponds to a probability value of 1 and is usually equal to the max response val, used for the response measure in Section (4.2), ϵ defines the tolerated difference of the warped neighbors and $\omega_{warpedDFD}$ is the damping weight. This measure mainly affects areas of complex motion in which the ME algorithm fails to correctly compute motion vectors. To extend the regions that are desired to be damped down a box filter is applied.

4.4.6 Moving edge detection measure

As already reported in several related work, FA are likely to occur close to sharp, moving edges. Chong and Krishnan [36] already introduced a moving edge detector into their MRF model and it appeared to be an effective technique in order to regularize false detections. When warping the neighboring images to the center image, pixels are translated by following the appropriate motion vector. Motion



(a) frame n - 1 **(b)** frame n **(c)** frame n + 1

(d) Result using TV-L1 optical flow **(e)** Result using Pixel motion flow

Figure 4.15: Probability measure based on DFD of warped neighbors, illustrated for images sequence *Dance2*. ME errors are mainly affected by these measure, which often occur in textured areas.

vectors do not always point to an exact coordinate, thus, pixels are bi-linearly interpolated. The bi-linear interpolation schema and a simple nearest neighbor interpolation method, for example, usually produce different results. In homogeneous regions the difference is not recognizable, but it is around edges. When calculating the difference between two different interpolation modes, e.g. between bilinear and nearest-neighbor interpolation, high values are typically measured on moving edges, but it also take effect at noisy areas. We use

(a) frame n - 1 **(b)** frame n **(c)** frame n + 1

(d) Result using TV-L1 optical flow **(e)** Result using Pixel motion

Figure 4.16: Probability measure obtained by moving edge detection, illustrated for image sequence *Surprise*. This measure affects only moving edges, which are generally likely to produce FA, other edges, such as dust spot boundaries are not affected.

$$m_m(z) = |C_n(z) - C_n[z + \vec{m}_z + 0.5]|, \tag{4.19}$$

where C_n is the warped neighbor, $\vec{m}_z$ is the motion vector estimated for pixel z of the center frame and [] indicate floor brackets. A probability measure is then inferred, which indicates the likelihood for each pixel that it lies on a moving edge:

$$p_{movEdges}(z) = \omega_{movEdges} \cdot \min\{\frac{\max\{m_m(z) - \epsilon, 0\}}{m_{m_{max}} - \epsilon}, 1\}, \tag{4.20}$$

where $m_{m_{max}}$ is the maximum difference which corresponds to a probability value of 1. ϵ defines the tolerated difference between the results of the two different warping methods and $\omega_{movEdges}$ is the damping weight. The novelty of our proposed moving edge detector is due to its ability to find exclusively moving edges. Static edges or dust spot edges are not affected. In Figure (4.16) the inferred probabilities are visualized.

4.4.7 Tracing of occluded/uncovered motion vectors

Detection of occlusion and uncovering areas can also be performed directly, since the motion boundaries and the motion directions are available.

The divergence measure is used to obtain motion boundaries and each motion vector which is measured at these boundaries is used to calculate the angle of the vector. When following each of the motion vectors the intensity of all passing pixels of the non-motion-compensated neighboring frame is compared with the intensity value of the starting position. If the difference does not exceed ϵ, we assume that there is neither an occlusion nor an uncovering area. If the intensity varies more than ϵ, we calculate the probability of each pixel belonging to an occlusion or uncovering area as follows:

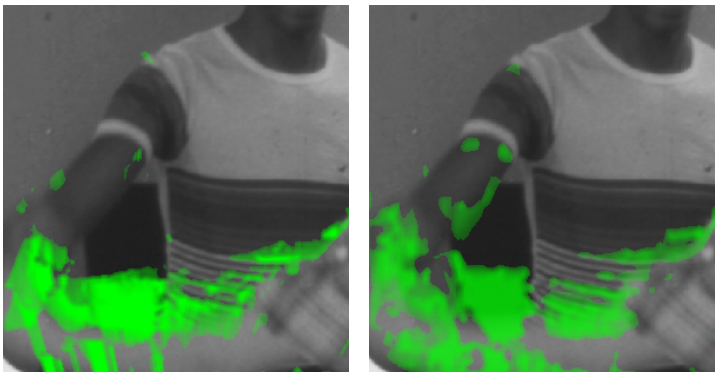
Let $\vec{m_z}$ be the motion vector estimated at position z and let z_{m_i} be the coordinates of its underlying elements. The coordinates are obtained by rotating a same-sized vector by the calculated angle. For the probability measure a d_{max} is defined, which indicate the max difference to which a value of 1 is assigned. A box filter is applied. Then the probability measure for this damping function is defined as

$$p_{occlusion}(z) = \omega_{occlusion} \cdot \min\{\frac{\max\{|I_{n\pm1}(z_{m_i}) - I(z)| - \epsilon, 0\}}{m_{o_{max}}}, 1\}, \tag{4.21}$$

where $\omega_{Occlusion}$ is the damping weight for the occlusion measure.

In Figure (4.17) the obtained probability values are visualized. For this measure we do not differentiate between occlusion and uncovering areas, as the probability should be equally distributed for both. The tracing of occluded/uncovered motion

(a) frame n - 1 **(b)** frame n **(c)** frame n + 1



(d) Result using TV-L1 optical flow **(e)** Result using Pixel motion

Figure 4.17: Probability measure obtained by detecting occlusions, illustrated by image sequence *Dance2*. The measure widely affects occlusion areas, but the affected region is less accurate than by other measures.

vectors produce an acceptable result when TV-L1 optical flow is used. Using Pixel motion for ME, the result is worse, as motion boundaries could not be detected satisfyingly by the divergence, see Figure (4.14e), and the motion vector directions do not seem to be reliable either.

(a) TV-L1 optical flow **(b)** Pixel motion

Figure 4.18: Response measured by the sROD based method considering the vertical neighbors.

4.5 Combined algorithms

In the last section different methods to obtain a response measure, a noise reduction operator and several damping functions are presented. This section shows a selection of the best methods, presented in this chapter. For each of the two ME algorithms a set of damping functions is chosen in order to produce the best results.

We are going to use the sROD_2 based response measure (considering the two vertical neighbors) and we apply the co-support operator in any case. In this way, a proper balance between noise reduction and preserving of fine structures, such as hairs, is given. In Figure (4.18) the response measure is illustrated, where incorrectly high measured responses around the hand are clearly visible. For this illustration, two vertical neighbors where considered for the sROD response measure. We also applied a penalty term for the sign constraint, as defined in Equation (4.2). For the visualization the max response value is set to 30.

FA should be damped down by the combination of an appropriate set of damping functions, which are discussed in the following subsections. Since the used ME algorithms are different in many ways, an individual set is assigned to each of the

methods.

After performing Equation (4.10) a combined probability measure is obtained, which is multiplied by the response in order to obtain a final response measure. Essentially, the final response R is a gray-level image whose non-zero values correspond to dirt particles with an associated likelihood of being part of dirt and dust. A final binary mask of dirt b_c can be obtained by thresholding R under a given likelihood $\eta \in [0, 1]$, where Λ is the number of gray values per channel. Using

$$b_c(z) = \begin{cases} 1, & \text{if } R(z) \geq (\Lambda - 1)\eta \\ 0 & \text{otherwise,} \end{cases} \tag{4.22}$$

the final binary dust mask is obtained, which can be used for the removal process.

4.5.1 Algorithm for TV-L1 optical flow

The TV-L1 optical flow offers a quite good ME. The motion boundaries are clearly measured and the motion field is overall smooth. If choosing this ME method the following damping functions work best:

All damping functions are blurred by a box filter with a kernel size of $\approx 1\%$ of the image with.

1. Density of motion vector field: As seen in Figure (4.12), occlusion and uncovering areas are completely detected, we can clearly recognize the motion during the 3 frame sequence. Since pixels, which lie in occlusion/uncovering areas, do not have references in both neighbors, high response values in those areas are not reliably classified as blotch pixels, thus those values can be damped down justifiably. As the density measure provides quit high values, the damping weight $\omega_{density}$ is slightly decreased.

2. Divergence of motion vector field: The divergence measure results in high values at motion boundaries. In Figure (4.14d) complex motion, in particular, is clearly detected. Similar to the density measure, the damping weight for the divergence ω_{div} is slightly decreased.

(a) Response measure **(b)** combined damping function

Figure 4.19: Combined probabilities of all used damping functions, where TV-L1 optical flow was used for ME. Those areas, that are due to be damped down are clearly recognizable by the yellow shade.

3. Difference of warped neighbors: Areas, which manifests a significant difference between the two motion compensated neighbors, are damped down, as those areas are obviously wrongly estimated. As illustrated in Figure (4.15), high differences were detected in areas of fine structures (stripes of T-shirt). Remaining high response values in such fine structured areas, increase the risk of introducing the construction of annoying artifacts in the removal process. The same constrained is used in the SDIjrs [21] method and is responsible for an improved performance when compared to the SDIa and (in most cases) SDIp methods.

4. Moving edge detection measure: Finally, the moving edge detector damps down all response values on and close to moving edges. This is necessary, since response values are likely to occur close to sharp edges due to inaccuracy of the ME algorithm. Even if the motion vector only varies slightly, the response measure might deviate heavily.

Each of the used damping functions was illustrated separately in Section (4.4). The combined set of used damping function is overlaid with the response measure and illustrated in Figure (4.19). The response measure is encoded in the red channel

and the damping probability in the green channel. Consequently, all response values which are going to be damped down are shown in a shade of yellow. The combined damping function is shown in Figure (4.19b), which has to be reversed before it can be multiplied by the response measure.

4.5.2 Algorithm for Pixel motion

The Pixel motion algorithm is less accurate than the TV-L1 optical flow. It is less smooth, has more outliers of motion vectors in homogeneous regions and cannot deal with illumination changes. However, the best set of damping functions appeared to be the following:

1. Difference of motion compensated neighbor: High differences of the motion compensated neighbors indicate unreliable areas and response values are damped down. Similar to the result produced when TV-L1 optical flow is used, fine structured areas are affected, as illustrated in Figure (4.15e). This measure is applicable for both ME algorithms, since no significant differences of the results are recognized.

2. Moving edge detection measure: Similar to the previous measure, where the motion compensated neighbors are compared, the moving edge detection measure seems to be independent of the ME algorithms. Both results accurately show the moving edges and consequently damps down occurring high responses.

3. Back-matching of motion vectors: The back-matching function, shown in Figure (4.13e), is not an accurate measure for occlusion, uncovering and complex motion, but generally, it takes effect in the right areas. This measure is only applicable for the automatic mode, since the ME is required in both directions, which will result in a higher computational cost.

4. Density of the motion field: As seen in Figure (4.12e), occlusion and uncovering areas are reliably detected. But, in contrast to the result when using TV-L1 optical flow, reliable areas might be marked as unreliable as well. Particularly, this effect occurs in presence of illumination changes, which is

(a) Response measure **(b)** combined damping function

Figure 4.20: Combined probabilities of all used damping functions, where Pixel motion was used for the ME. Those areas, that are due to be damped down are clearly recognizable by the yellow shade.

due to a limitation of the used ME algorithm. Therefore smaller probability values are ignored to counteract this effect.

Each of the used damping function is illustrated separately in Section (4.4). The combination of all damping functions and the resulting effect on the response measure is illustrated in Figure (4.20).

4.6 Discussion

In this chapter we have described the proposed detection approach single-frame-defect detection approach based on the Reduction of FA by analysing the estimated Motion followed by an Extraction of Damping functions (rFAMED) in detail. The probability based response measure was chosen to keep essential information, that was required later on in order to reduce FA. For the response measure a sROD based detection measure was chosen, since it is fast and simple and good results were reported in related work. The co-support operator provides an effective method in order to avoid single outliers due to noise and to increase responses within and close to dust spots. In addition, we have presented several damping

functions in order to affect areas of occlusion, uncovering and complex motion, where FA due to ME errors are likely to occur. Those damping measures were obtained by conducting an comprehensive analysis of the computed motion field and the appropriate warped neighbors. The whole approach is widely independent of the used ME algorithm, thus any algorithm can be applied. The proposed measures cover all precarious areas of an input frame that are likely to result in false detections. The aim is to damp down high responses occurring in such areas of PM in order to significantly reduce false detections which endanger artifacts in the restored frame. In the next chapter experiments and evaluations are accomplished to illustrate the performance of rFAMED, our novel detection method.

Chapter 5

Experimental Results and Evaluation

5.1 Introduction

In this chapter a comparative evaluation of the performance of the proposed novel dust and dirt detection algorithm is carried out. We show the results and the improvement over previous detectors on several different image sequences, including material with noise, complex motion, occlusions and uncovering areas. In order to compare the detectors against each other, a quantitative evaluation schema and an objective ground truth is chosen. Since most detectors require predefined thresholds and different parameters, we use ROC curves for comparative illustrations.

The proposed algorithm is compared against the single-frame-defect detector integrated in the DIAMANT film restoration software, which we aim to outperform with this work. Furthermore, it is compared against five baseline detectors which are well known in the literature and which were already discussed in Section (2.2). In the following, the evaluation methodology is explained in detail, in particular, we focus on ROC analysis in order to evaluate and compare the different single-frame-defect detectors. Furthermore, we will discuss issues and limitations arising from the available ground truth data (GT). Finally, the performance is evaluated quantitatively as well as qualitatively, followed by a detailed discussion of the obtained results.

5.2 Evaluation methodology

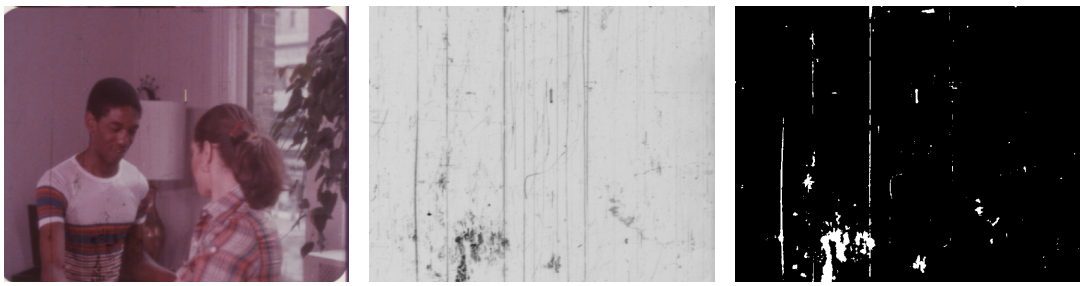
To afford a comparable, statistical evaluation, a representative evaluation schema is carried out. Commonly, the well known ROC curve [67] approach is used for evaluating a detector's performance. The performance is dependent on both the correctly detected parts and the false detections. Since a high accuracy is worthless as long as the FA are high as well (following [5] a FAR $< 0.1\%$ is aspired), the right balance between accuracy and FA has to be found. Thus, the well known ROC curve [67] is an applicable approach, since it shows the CDR against the FAR. For the evaluation methodology an objective GT is necessary in order to compute these values. As a corresponding dirt free image sequence is usually not available, another objective measure is required. In related work regarding dirt and dust detection algorithms, e.g. in [5], [14], etc., infrared (IR) scans have been established as objective GT. Furthermore, artificially corrupted image sequences were often used in related approaches, e.g. in [10], [27], etc., where the corresponding GT was inferred from the artificially corrupted test data. Similarly, e.g. in [40], real blotches were added to the test data by copying and pasting of previously detected ones. Although good results have been reported with these artificially added degradations, they actually do not reflect the real situation and hence it does not represent a reliable evaluation. Consequently, we have chosen to use IR scans as objective GT, since they are inferred from real corrupted image sequences and reliably show defects, that are physically present on the film.

5.2.1 Ground truth (GT)

For the comparative evaluation of the detector's performances we have chosen IR scans, made available from INA (Institut National de L'Audiovisuel)[1] as objective GT. Although, IR scans can yield limitations when using them as GT in a pixel based evaluation, they allow a quantitative comparison of different detection algorithms. Possible limitations are discussed later in this chapter.

GT masks are obtained by thresholding IR scans of archived films under an appropriate value. An IR scan is a gray level image, as shown in Figure (5.1b), in

[1] http://www.ina.fr/



(a) corrupted image data **(b)** corresponding IR scan **(c)** GT mask

Figure 5.1: GT for frame 9 of image sequence *Dance2*. The IR scan in (b) is thresholded by a value of 170, where values smaller than 170 are classified as corrupted pixels (displayed in white) and values greater or equal are classified as clean (displayed in black) in the resulting binary GT mask (c).

which all physical defects are contained. Due to their non-transparency, dust, dirt and other defected areas are detectable by an infrared sensor and always appear darker than uncorrupted regions. The darker a defected pixel is represented, the higher is the probability that it refers to a pixel of dirt [5]. The thresholded GT maps are binary images, displaying dirt pixels as white and clean pixels as black, illustrated in Figure (5.1). The GT mask is then used to calculate the CDR and FAR, which are required for the ROC analysis.

5.2.2 Receiver operating characteristics (ROC)

The ROC curve approach provides a comprehensible way in order to assess the accuracy of predictions. We are going to use this schema to predict a binary outcome, i.e. the current pixel belongs to a defect region or it does not. Thus, a pixel based evaluation is performed, where each pixel belongs to one of the four classes, listed in Table (5.1).

The ROC curve is a graphical illustration in order to explore the trade-offs between correct and false detections. The traditional ROC curve plots sensitivity (true positive rate or CDR) on the vertical axis and the specificity (true negative rate or FAR) on the horizontal axis. A quantitative variable is needed to guide the decision [67]. In this evaluation, the threshold, which is used to obtain the final

	defected pixel ($t_p + f_n$)	**not defected pixel** ($f_p + t_n$)
detector classifies pixel as defected ($t_p + f_p$)	true positive (t_p)	false positive (f_p)
detector classifies pixel as not defected ($f_n + t_n$)	false negative (f_n)	true negative (t_n)

Table 5.1: Frequency distribution of the variables for the ROC analysis, required to compute CDR and FAR.

binary dust mask, is chosen to be the quantitative variable. For each threshold the CDR and the FAR are calculated by using

$$CDR = \frac{t_p}{(t_p + f_p)} \tag{5.1}$$

and

$$FAR = \frac{f_p}{(f_p + f_n)}, \tag{5.2}$$

where t_p are true positives, f_p are false positives and f_n represent false negative elements, further explained in Table (5.1). Thus, each point of the ROC curve corresponds to a specific threshold. Connecting those points leads to the so-called empirical ROC curve.

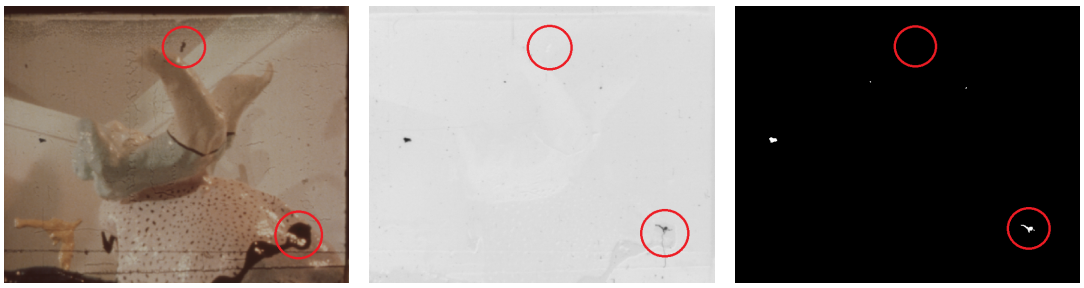
Since the measures for correct and false detections are pixel based, Equations (5.1) - (5.2) can be rewritten as

$$CDR = \frac{Count(D_{GT} \& D_{DM})}{Count(D_{DM})} \tag{5.3}$$

and

$$FAR = \frac{Count(\overline{D_{GT}} \& D_{DM})}{Count(\overline{D_{GT}})}, \tag{5.4}$$

where *Count* is a function which defines the number of set pixels in the appropriate binary mask, D_{GT} and D_{DM} are the two required binary masks, which are the GT



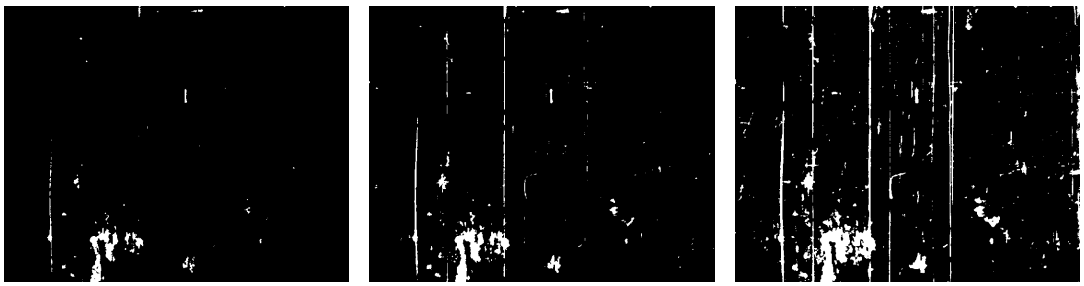
(a) corrupted image **(b)** corresponding IR scan **(c)** resulting GT mask

Figure 5.2: Problematic IR scan as GT. Firstly, there is a dust spot in the original image, which will contribute to the FA, since it is not present in the appropriate GT mask. Secondly, a dust spot that is located on the IR scan cannot be found by any detector, since the dust spot is hidden due to the image content.

and the computed dust mask, respectively. The &-Operator is the logical AND, in other words all pixels that are set in both masks, and $\overline{D_{GT}}$ specifies the complement of the GT, consequently "the non-set elements".

5.2.3 Limitations

Using IR scans as GT has several limitations for the dirt and dust detection performance assessment. In the first place, IR scans are not always available, since a special scanner is necessary to produce an additional IR scan for every image frame. Furthermore, they can only be made on color films, since the silver image in monochrome film is opaque to infrared radiation [52]. Thus, the evaluation often can only be done for an exclusive set of image sequences. Secondly, IR scans typically do not contain solely single-frame-defects, but also other defects like line scratches. Line scratches usually occur at the same position in several consecutive frames, thus single-frame-defect detectors are not able to detect them, which will always lead to a decreased CDR. Furthermore, defects on a film might be already copied. IR scans only catch blotches, if they are physically located on a film reel. Copied dust will not appear on the GT mask and will bias the result. Finally, the result heavily depends on the chosen threshold for the IR scan, since dust pixel only appear in the GT mask if they exceed the appropriate threshold.



(a) Threshold 150 **(b)** Threshold 170 **(c)** Threshold 190

Figure 5.3: GT for image sequence *Dance2* thresholded at different values.

Choosing an appropriate threshold for the infrared scan arises to be very sensitive to the performance: a lower threshold results in a less sensitive GT and increases the CDR while decreasing the FAR. Using a higher threshold results in a GT containing more low contrasted and fine structured dust spots and will decrease the overall detectors performance. This can be explained by the used formula for calculating CDR and FAR defined in Equation (5.1) and (5.2). In Figure (5.3) three binary masks of dirt are illustrated, where the IR scan from Figure (5.1) was thresholded by three different values. When using a higher threshold, the content of the shown GT mask is not limited to single-frame-defects only, also line scratches are included. When using a lower threshold, line scratches are widely excluded, but so are fine structured dust spots, e.g. hairs. Those pixels are contributing to the FAR in the evaluation curve, even though all of them were correctly classified as defected. Since the measure is pixel-based, the fact, that dust spots have a smooth border, can also influence the result. When using a higher threshold, the border pixels may just fall under the threshold and thus dust spots are displaced weaker. In contrast, a lower threshold forces dust spots to be more extended and stronger. This does not impact the result if only small-sized dust spots are contained in the center frame, but it surely does, in case of larger dust spots, since only the exact overlap of dust spots in the GT and those contained in the computed dust mask will exclusively increase the CDR and decrease the FAR.

In Figure (5.4) the impact on the resulting ROC curve is illustrated. In the literature it was argued, that the threshold was defined to be 170 in order to

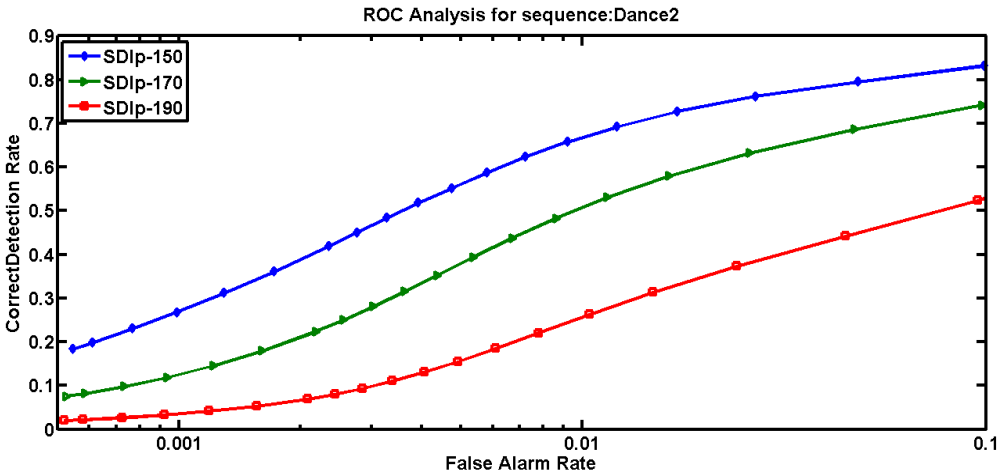


Figure 5.4: Difficulty of choosing the right threshold for IR scans. This curves were produced by the SDIp detector for the sequence *Dance2*. The blue curve represents the result when using a GT thresholded by 150, for the green one a threshold of 170 and for the red curve a threshold of 190 was used.

ignore the effect of scratches and other semi-transparent artifacts while keeping dirt particles. Thus, choosing a threshold of 170 seems to be suitable.

Since IR scans appear to be problematic a qualitative evaluation of the novel detector will be performed in addition.

5.3 Test sequences

To cover the huge diversity of film material to be restored, a variety of test data is chosen. Today's film scanners produce images with a resolution ranging from SD to 8K, where the content differs in its level of noise and film grain, in the complexity of containing motion as well as in other quality aspects such as illumination changes or image contrast. Since IR scans are required in order to do an evaluation, the number of available test sequences is limited.

Among others, we have used six broadcast resolution (760 x 560) sequences, which were already used in the evaluation of several related work within the *PrestoSpace* project, e.g. in [5], [14]. In Table (5.2) the chosen image sequences are listed and described in detail.

Image sequence (# frames)	**Resolution**	**Image sequence content**
Lady and doll (293)	SD (720×576)	A woman next to a doll is moving her hand and her head: no complex motion, contains large dirt areas, camera zooming and shaking, contains local motion and partially textured back-ground.
Art (84)	SD (720×576)	Presentation of works of art, strong translational motion, high amount of small dust spots, no local motion.
Cigaret (269)	SD (720×576)	A group of original inhabitants, shot in the nature, smoking a cigarette: contains poor contrast, apparent camera shaking, non-rigid human motion, translational motion, strong textured background.
Surprise (264)	SD (720×576)	Walking lady, in background a dancing couple: no complex motion, middle sized dust spots, high amount of hairs.
Dance2 (205)	SD (720×576)	A dancing couple: very challenging sequence, slow camera motion, fast and complex object motion, textured background, severe scratches.
JazzMan (11)	2K (2048×1556)	A couple is walking through the streets, a man is playing saxophone on the sidewalk, marginal object motion, high level of fine noise.

Continued on Next Page...

Table 5.2 – Continued

Image sequence (# frames)	**Resolution**	**Image sequence content**
Stadt ohne Juden (26)	SD (720×576)	A waiter is serving drinks: fast and complex motion, illumination changes, high level of coarse film grain.
Elvis2 (77)	SD (720×576)	Elvis is blowing someone a kiss, partially complex motion, illumination changes, low amount of defects.
Hannes runs (115)	HD (1280×720)	A person is running through the park: strong and complex, local object motion, new film material, does not contain any dirt or dust.
Cat (115)	HD (1280×720)	A cat in the meadow is shaking its head: strong and complex, local object motion, new film material, no defects contained.

Table 5.2: Image sequences used for the evaluation

5.4 Results

The proposed dirt and dust detection algorithm is compared against well known single-frame-defect detection algorithms described in Section (2.2) and the previously integrated detector of the DIAMANT film restoration software.

Spatial filtering methods and methods without motion compensation are not included in our evaluation, since in several related work, e.g. in [5], it was confirmed that those methods produce worse results. MRF was not implemented either,

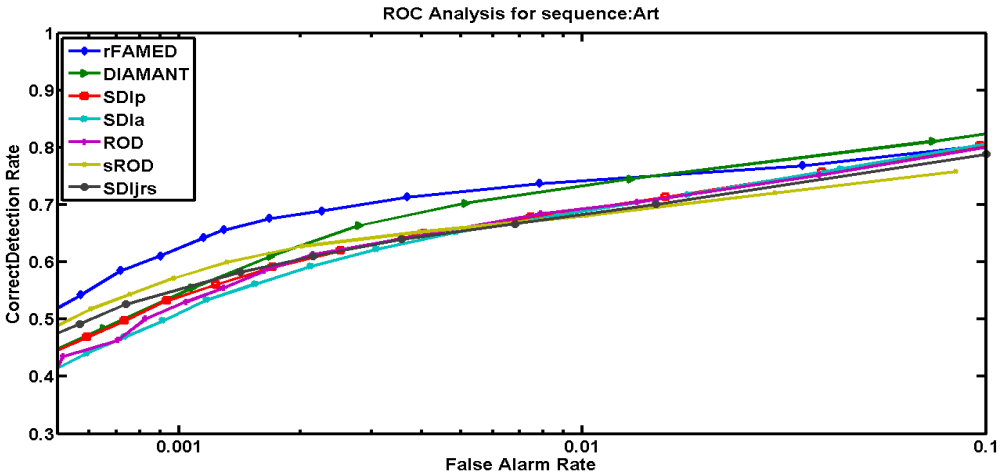
since a high computational cost is not applicable for the practical application of automatic film restoration. Consequently, the following evaluation contains the performance results of seven single-frame-defect detectors, which are SDIa, SDIp, SDIjrs, ROD, sROD, DIAMANT and our novel detector, denoted as rFAMED.

Since we have two ME algorithms at our disposal, this further aspect is also considered and discussed. First, we will present the quantitative results by using the ROC technique, followed by a qualitative evaluation to exclude possible misleading results due to the chosen GT for the quantitative evaluation technique.

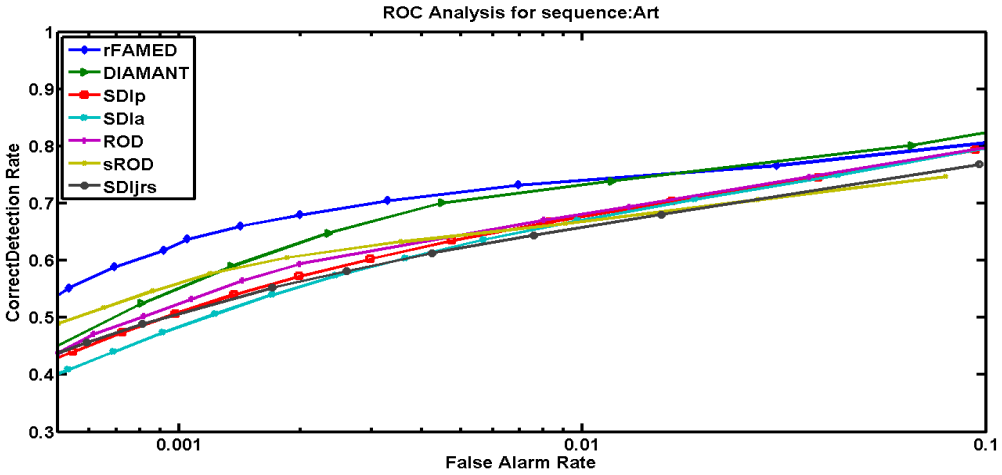
5.4.1 Quantitative evaluation

In total, seven ROC curves are calculated for each sequence. To ensure fair performance comparisons, we have used the same ME algorithm with an equal configuration of its parameters for each evaluation run. Furthermore, one set of thresholds was chosen for each detection algorithm, which was adopted as it stands for each image sequence. Figures (5.5 - 5.10) illustrates the results of each examined detector for the image sequences, listed and described in Section (5.3). Note, that the x-axis represent the FAR in a logarithmic scale and the y-axis shows the CDR, both expressed in percentages. For a more comprehensible visualization every chart only shows the most interesting parts of the curves, which in particular is the part between 0.001 and 0.01% of FA, since a very high CDR is useless at the expense of a higher FAR. We do not apply any preprocessing, such as deflicker or image stabilization, which obviously would lead to a better performance of all examined approaches.

The chart in Figure (5.5) shows the results for the sequence *Art*, which does not include local motion, but a strong translational motion. For this sequence, the results are widely independent of the used motion algorithms, since the curves in both figures are similar. rFAMED outperforms all other state of the art detectors, in particular when applying a higher threshold (arranged in the left part of the figure). Since there is neither complex motion nor noise contained, all examined detectors result in a similar performance. The superiority of our detector can be mainly explained by the co-support-operator, since it excludes single outliers and



(a) Using TV-L1 optical flow

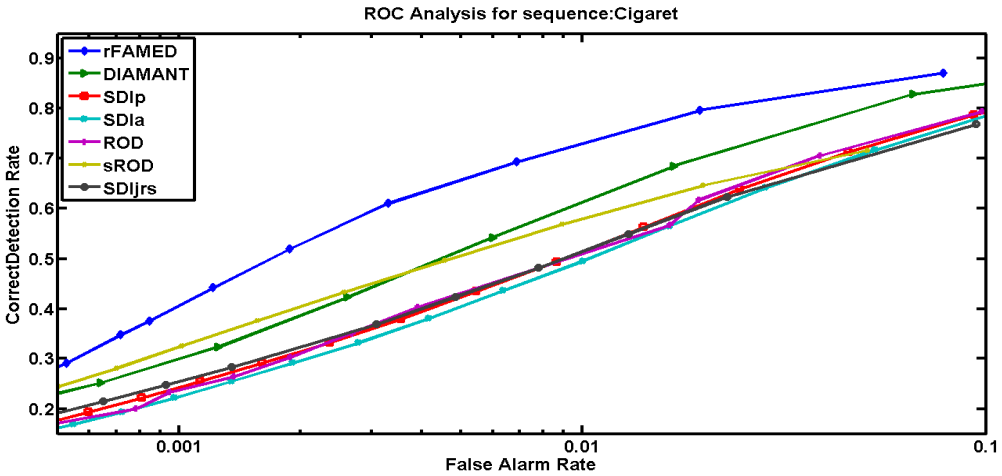


(b) Using Pixel motion

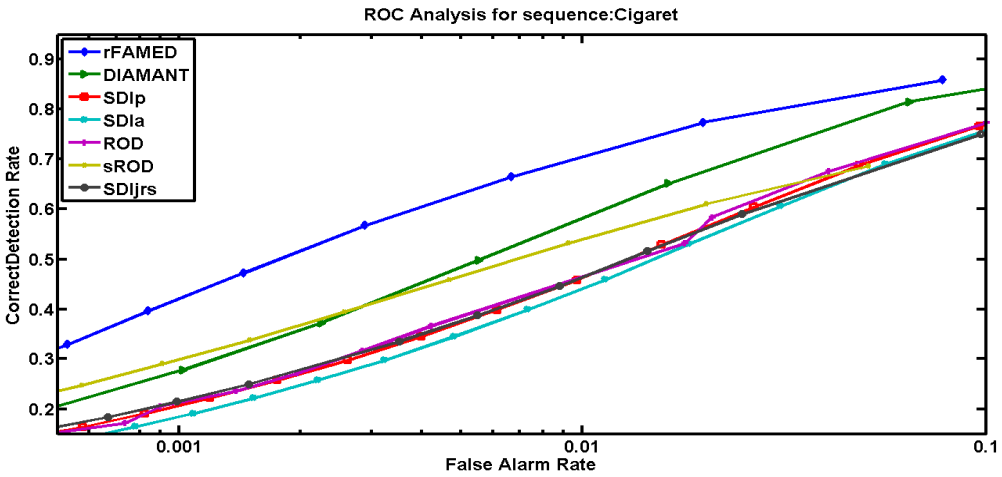
Figure 5.5: ROC analysis of image sequence *Art*.

expands dust spots, which are possibly not detected completely.

The *Cigaret* sequence is challenging due to its structured background and the resulting low contrast between background and dust spots. In Figure (5.6) the results obtained for this image sequence are illustrated. We can notice an overall lower CDR and simultaneously a higher FAR. This seems to be caused by the low frequency and the low visibility of blotches and the textured background that generates many FA. Additionally, this sequence contains camera shake. The moving



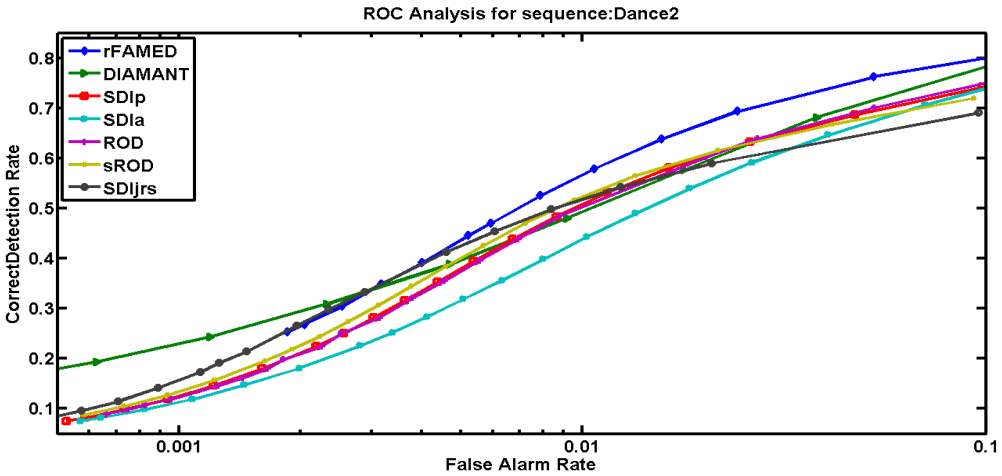
(a) Using TV-L1 optical flow



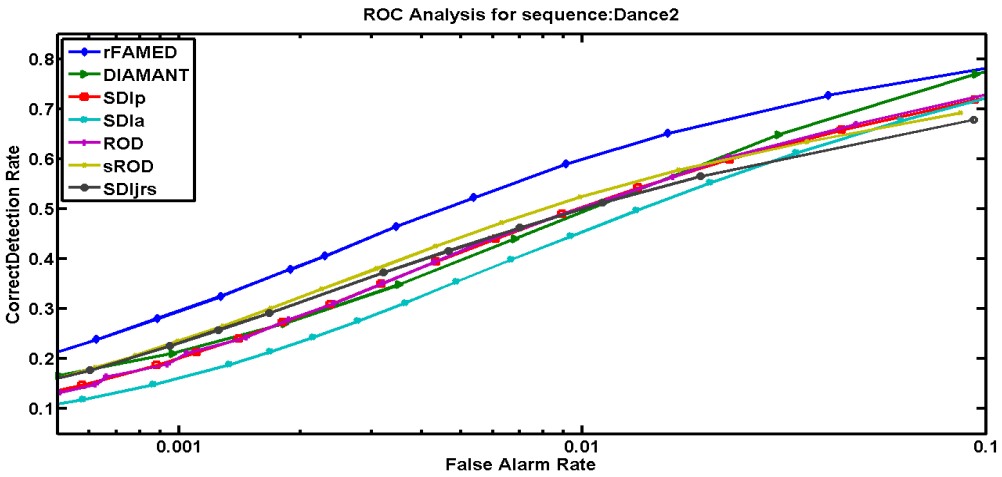
(b) Using Pixel motion

Figure 5.6: ROC analysis of image sequence *Cigaret*.

trees and leaves in the background seem to be difficult for both, the temporal and the spatio-temporal detectors. The large gap between the novel detector and the other techniques can be explained due to the fact, that a lower threshold can be applied, since false detections can be avoided in a further step. Applying the co-support operator and the moving edge measure further improves the result. Thus, a lower threshold facilitates finding also low contrasted blotches. As reflected by the result obtained for ROD and also reported in [5] more spatial support usually



(a) Using TV-L1 optical flow



(b) Using Pixel motion

Figure 5.7: ROC analysis of image sequence *Dance2*.

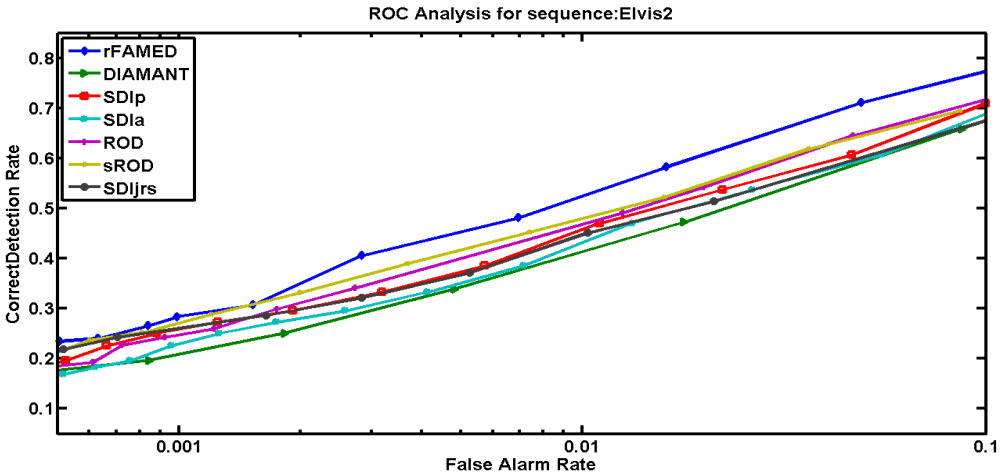
leads to a higher FAR in frames with textured background. Even though the sROD detector includes spatial information, it has a better performance than the other state of art methods. The advantage of sROD over ROD is a lower strictness, by using only one threshold. Again, quite similar results are produced independent of the used ME algorithm.

In Figure (5.7) the ROC analysis of the *Dance2* sequence is presented. The sequence has a very challenging content, i.e. very complex motion and consequently

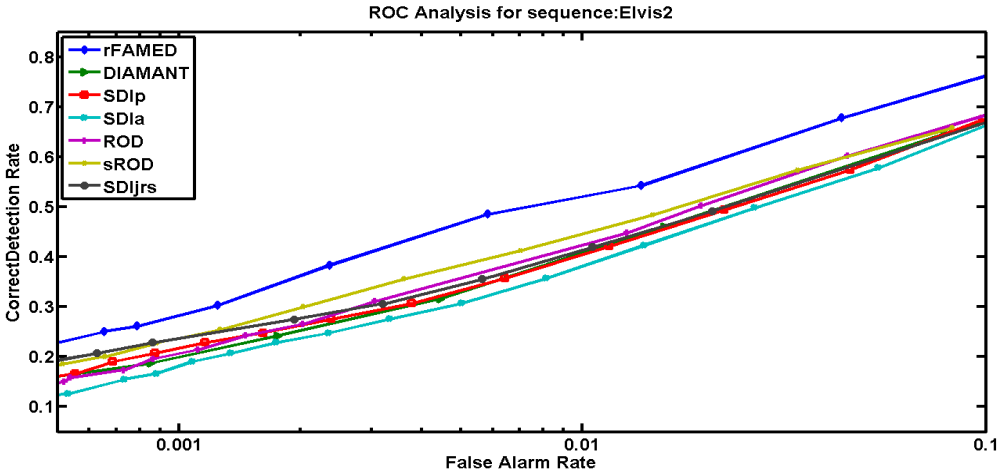
severe occlusion and uncovering areas. In general, our novel dirt and dust detector is configured for lower thresholds, since it includes FA reduction measures. At the left part of the shown figure, rFAMED is equal or worse than other techniques, since low contrasted and fine structured dust falls under the threshold and is not detected when using a higher threshold. In this part, the performance of the DIAMANT is better, since it uses an adaptive dilation dependent on the size of the appropriate dust spot. It further includes some techniques for excluding dust if the replacement does not match with the surrounding background. Since the motion is very complex, the damping functions should be configured very strict for such a sequence, which means, that the tolerances should be set smaller and the damping at the appropriate areas should be stronger, otherwise FA are left due to occlusions, uncovering areas, motion blur, moving edges, etc. Since the *Dance2* sequence contains a severe amount of line scratches, which are contained in the GT mask, but which does not belong to single-frame-defects and consequently cannot be detected by such a detector, the CDR values in the ROC curve are under-estimated and the FAR values over-estimated.

When using Pixel motion, the proposed detector turns out to be better than DIAMANT, since it additionally focus on ME errors, by applying stricter damping. The presented curves indicate an overall lower amount of CD, which again happens due to a higher amount of FA. The low CDR in this case can be explained by the persistence of scratches (which cannot be found by a single-frame-defect detector) and by the complexity of the contained object motion.

In Figure (5.8) the results for a scene showing Elvis on a balcony are illustrated. Only a few defects are contained, but camera shake and illumination changes are present. Due to illumination changes all the investigated single-frame-defect detection algorithms face some further challenges and result in a higher FAR. The IR scans do include dust spots, but in a very low contrast, thus the GT masks barely contain any defects. Consequently, a low number of frames are contributing to the shown results, since GT masks, that contain only zero elements, are excluded. In the *Elvis2* sequence, the consequences of the problematic GT appear in a large extent. Many low contrasted dust spots, which are found by the detectors are not included in the GT. In the region of local motion (moving hand for blowing a kiss), all examined techniques produce a high amount of FA.



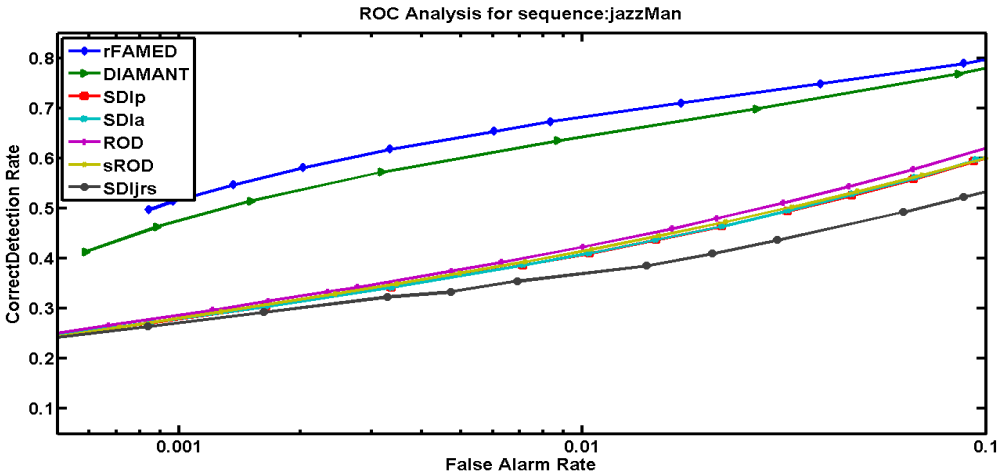
(a) Using TV-L1 optical flow



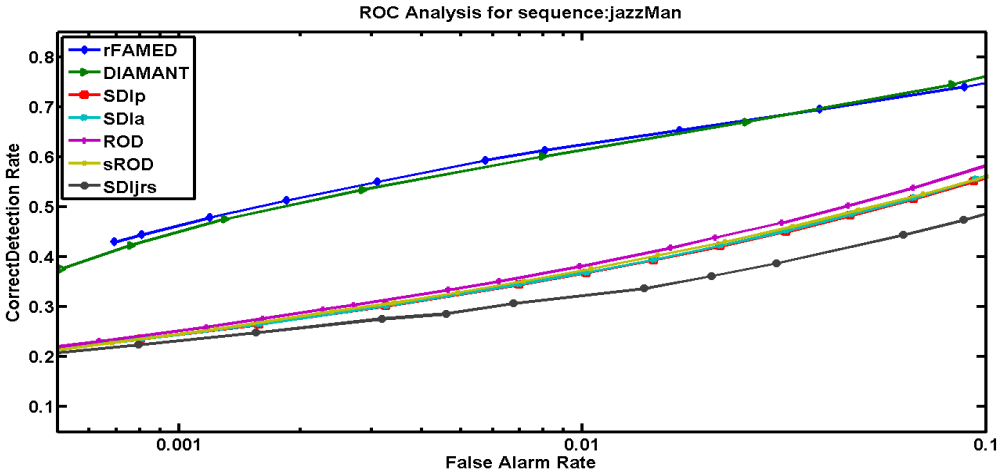
(b) Using Pixel motion

Figure 5.8: ROC analysis of image sequence *Elvis2*.

In contrast, rFAMED is able to significantly reduce FA in those areas of such complex motion. The temporal detectors, SDIa and SDIp in particular, have problems at moving edges. sROD is much better in this situation. The DIAMANT detector produces many FA due to complex motion as well. The shown ROC curve does not reflect the real performance of the novel detector. In Section (5.4.2) the image sequence is also discussed in a qualitative analysis. The density measure is effectively employed, which means that the right areas are damped down. When



(a) Using TV-L1 optical flow



(b) Using Pixel motion

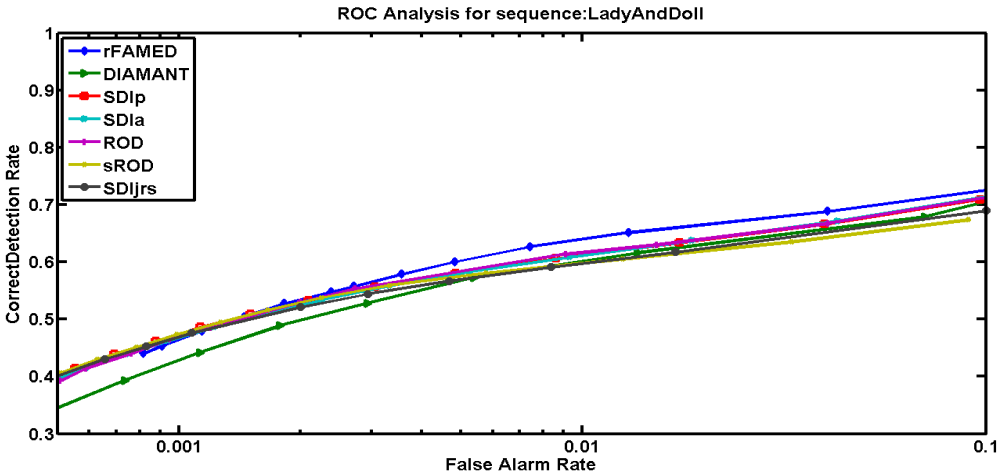
Figure 5.9: ROC analysis of image sequence *JazzMan*.

Pixel motion is used for the motion compensation, rFAMED is performing even better. This observation can be traced back to the fact, that the damping of the density is applied to a stronger extent when Pixel motion is used. SDIa, SDIp and SDIjrs produce much more FA when testing this sequence. In the case of TV-L1 optical flow, the DIAMANT detector is performing the worst. This is a surprising outcome, since it usually performs better than the baseline detectors. We can argue, that this result was established due to the strong dilation, which is

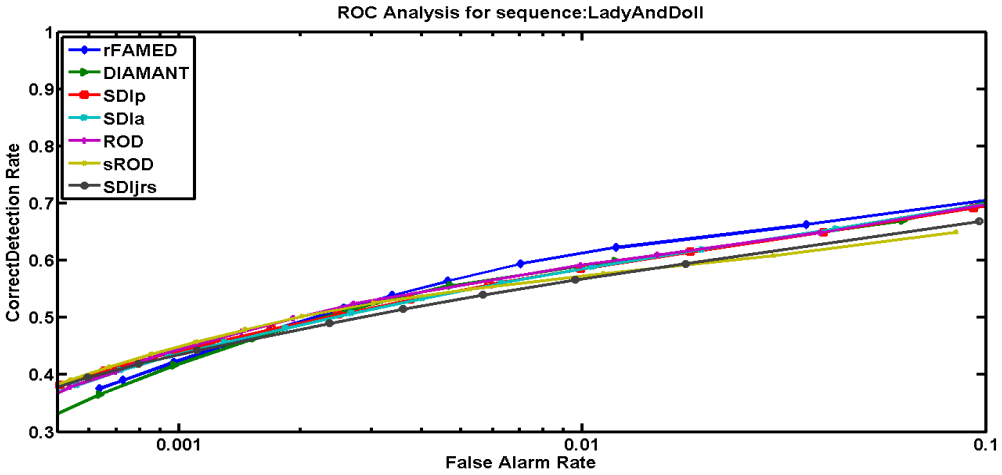
also performed on false detections.

In Figure (5.9) the results of the examined detectors for the *JazzMan* sequence are evaluated. The *JazzMan* sequence is scanned in 2K and contains a very high amount of fine noise. No challenging object motion or illumination changes are included. The temporal filtering methods show the worst results, since they do not include any spatial information and thus are not able to avoid false detections due to noise. The DIAMANT detector in contrast uses a protection technique which is able to avoid fine noise. The novel detector has the best performance due to the sROD based response measure and the additionally applied co-support operator. Furthermore, the moving edge detector slightly affects areas of noise and thus it contributes to the damping of responses in noisy areas. As expected, the spatio-temporal filtering methods are outperforming the temporal ones. Surprisingly, the SDIjrs technique result in a worse performance than SDIa, although SDIjrs turned out to produce much less FA, than the SDIa detector. The reason for the diverging curves is, that the CDR of SDIa is higher than that of SDIjrs.

The *Lady and Doll* sequence does not include challenging parts, such as complex motion or noise. Mainly translational motion and low object motion is contained. In Figure (5.10) the corresponding ROC analysis is visualized. There is no significant difference between any of the seven examined detection algorithms. All of them have a rather low detection rate, which can be explained by the following: The sequence includes several low contrast hairs and other fine structured dust spots, which cannot be detected by the examined techniques. In addition, many small-sized dust spots are present, which do not appear in the GT mask. The problem of the GT is clearly demonstrated: All techniques produce very good result when comparing the original image with the resulting dust mask. This is not reflected in the presented ROC curve, shown in Figure (5.10). All detectors produce FA in structured areas, i.e. the striped clothes of the lady. The temporal filtering methods, in particular, show false detections in the region where the lady is moving her hand.



(a) Using TV-L1 optical flow

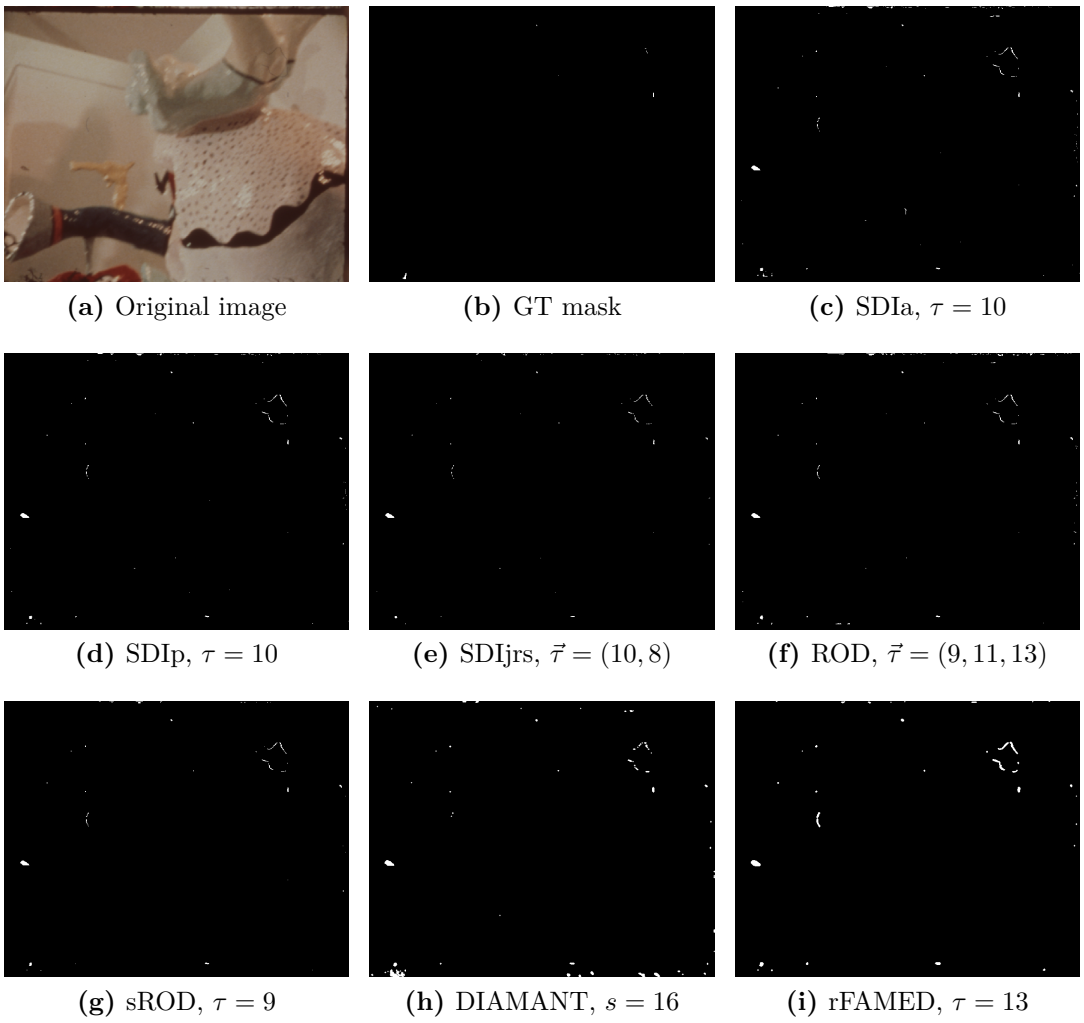


(b) Using Pixel motion

Figure 5.10: ROC analysis of image sequence *Lady and doll*.

5.4.2 Qualitative evaluation

In the following several selected detection results are presented qualitatively, in order to demonstrate for selected frames the effectiveness of the novel approach. For the results shown in Figures (5.11 - 5.17) the threshold (or the sensitivity in case of the DIAMANT detector) has been adapted for each sequence in order to produce a high quality result. For all following samples, we show the original and



(a) Original image **(b)** GT mask **(c)** SDIa, $\tau = 10$

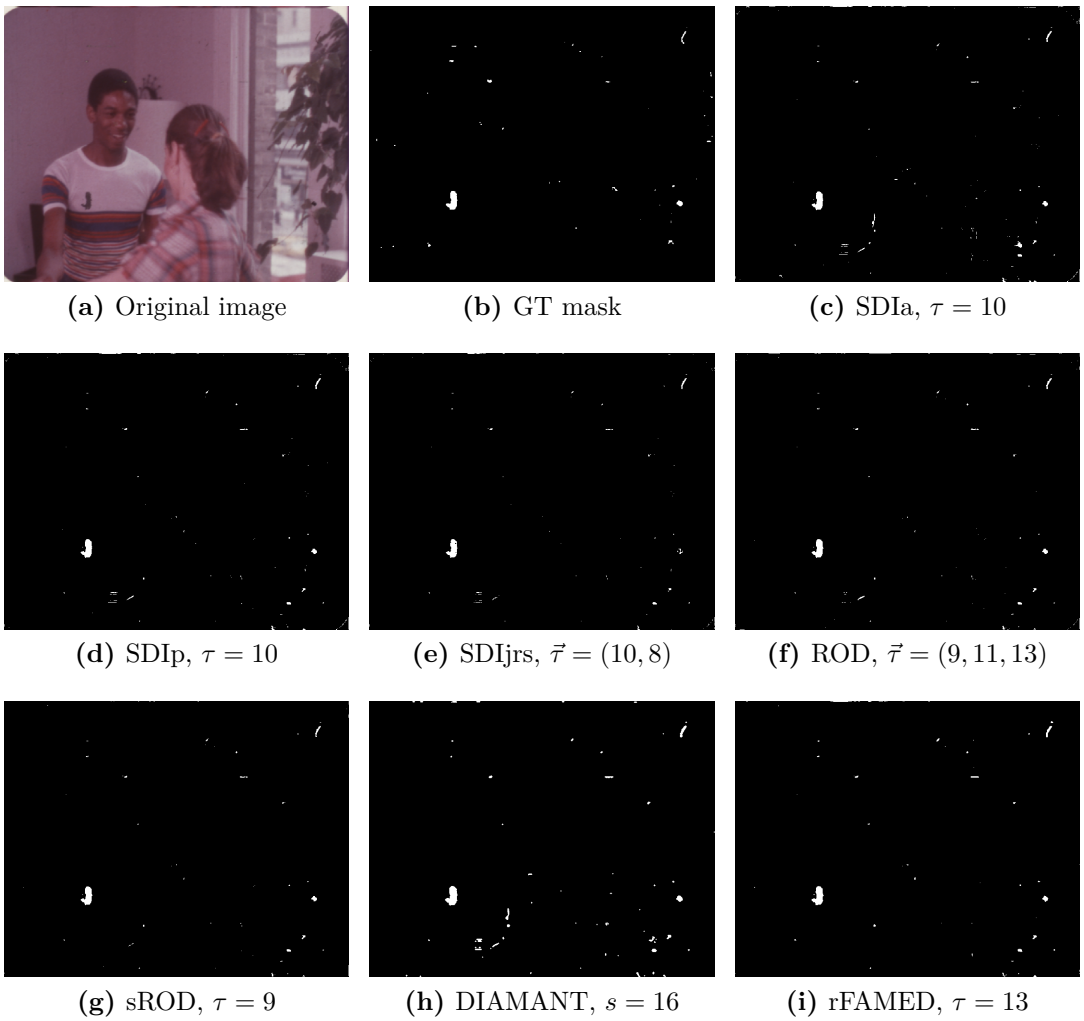
(d) SDIp, $\tau = 10$ **(e)** SDIjrs, $\vec{\tau} = (10, 8)$ **(f)** ROD, $\vec{\tau} = (9, 11, 13)$

(g) sROD, $\tau = 9$ **(h)** DIAMANT, $s = 16$ **(i)** rFAMED, $\tau = 13$

Figure 5.11: Qualitative performance analysis of frame 27 of sequence *Art*.

the computed dust masks by all examined detectors. Since the available GT does not always represent a reliable reference for the computed dust mask we provide the corresponding GT as well in the comparative, qualitative analysis.

In Figure (5.11) we show the resulting dust masks of frame 27 of image sequence *Art*. There are several mentionable facts regarding this frame. First, the GT is incomplete, therefore the corresponding CDR and FAR cannot be calculated correctly. The obtained dust masks are similar, but the novel dirt and dust detector



(a) Original image **(b)** GT mask **(c)** SDIa, $\tau = 10$

(d) SDIp, $\tau = 10$ **(e)** SDIjrs, $\vec{\tau} = (10, 8)$ **(f)** ROD, $\vec{\tau} = (9, 11, 13)$

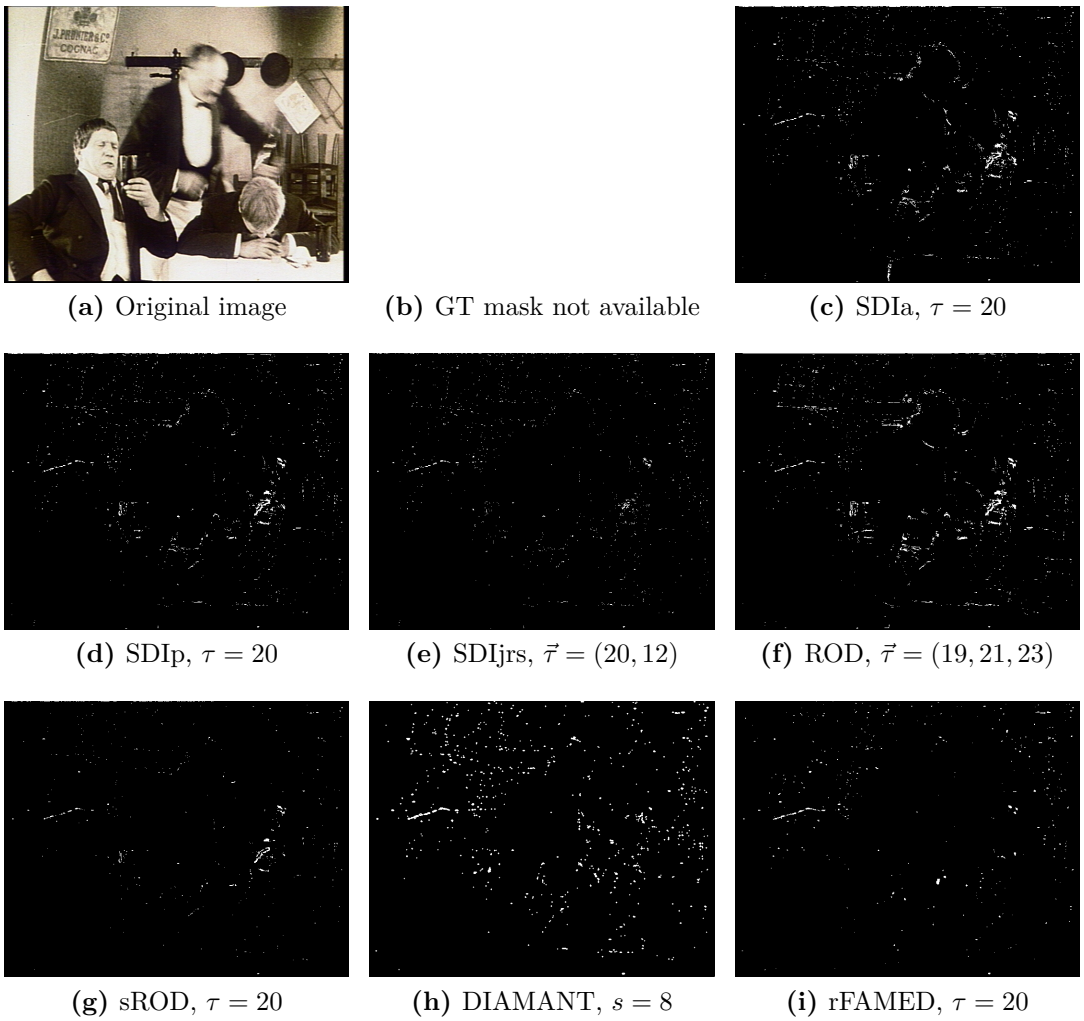
(g) sROD, $\tau = 9$ **(h)** DIAMANT, $s = 16$ **(i)** rFAMED, $\tau = 13$

Figure 5.12: Qualitative performance analysis of frame 48 of sequence *Dance2*.

is able to detect the fine structures as well, such as the containing hair.

In Figure (5.12) the qualitative analysis of a frame chosen from sequence *Dance2* is illustrated. This sequence contains very complex motion, thus a high FAR can be expected. rFAMED has the lowest FAR, since the density and the moving edge measure target the occlusion, uncovering and complex motion areas and damp down the response in those parts.

For the sequence *Stadt ohne Juden*, a higher threshold results in a better per-



(a) Original image **(b)** GT mask not available **(c)** SDIa, $\tau = 20$

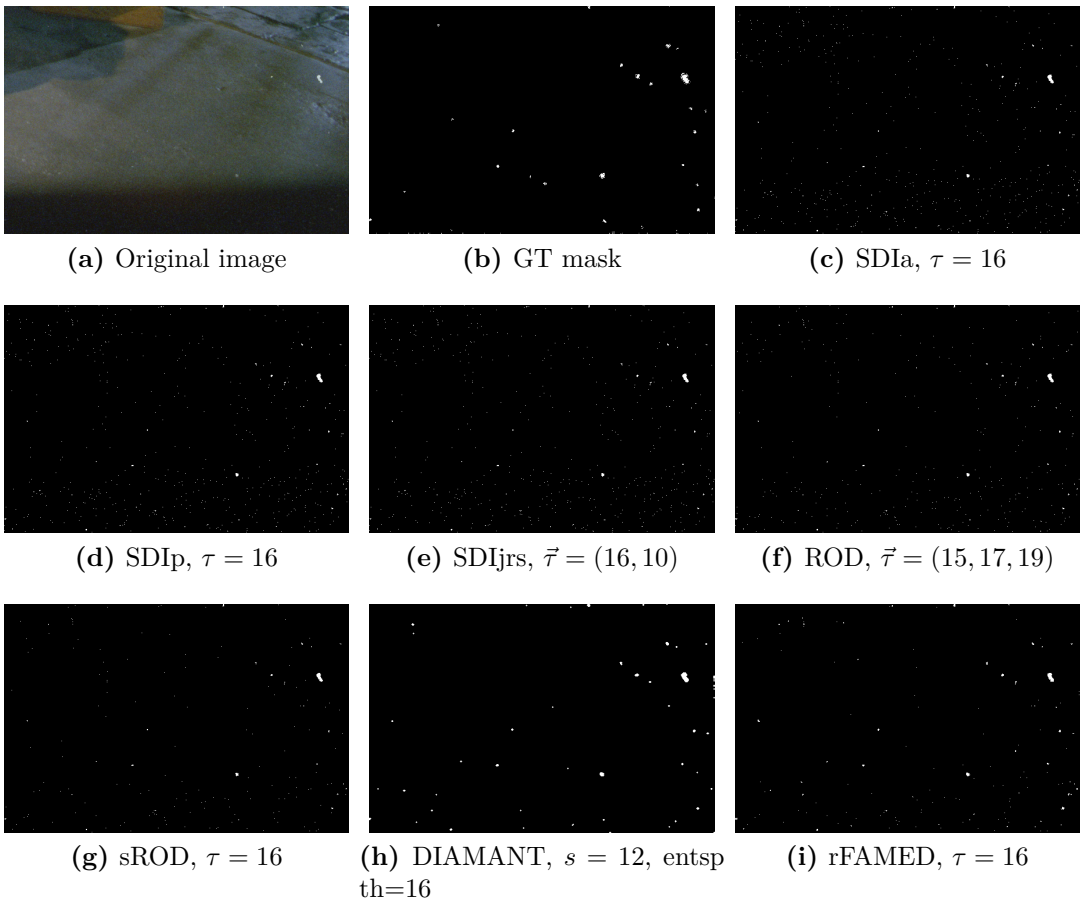
(d) SDIp, $\tau = 20$ **(e)** SDIjrs, $\vec{\tau} = (20, 12)$ **(f)** ROD, $\vec{\tau} = (19, 21, 23)$

(g) sROD, $\tau = 20$ **(h)** DIAMANT, $s = 8$ **(i)** rFAMED, $\tau = 20$

Figure 5.13: Qualitative performance analysis of frame 7 of sequence *Stadt ohne Juden*.

formance. Thus, we have adapted the thresholds appropriately, as listed in Figure (5.13), where the obtained dust mask for frame 7 are shown. Even when using a high threshold (or a low sensitivity in case of DIAMANT), FA due to noise or film grain persists. An even higher threshold would result in a lower CDR, which is not desirable.

In the next analysis, illustrated in Figure (5.14), frame 5 of sequence *JazzMan* was selected and a dust mask was calculated by each of the examined detectors.



(a) Original image **(b)** GT mask **(c)** SDIa, $\tau = 16$

(d) SDIp, $\tau = 16$ **(e)** SDIjrs, $\vec{\tau} = (16, 10)$ **(f)** ROD, $\vec{\tau} = (15, 17, 19)$

(g) sROD, $\tau = 16$ **(h)** DIAMANT, $s = 12$, entsp th=16 **(i)** rFAMED, $\tau = 16$

Figure 5.14: Qualitative performance analysis of frame 5 of sequence *Jazzman*.

Since a high level of fine noise is contained in this sequence, we have increased the thresholds. Although, we have adapted the threshold for every detector, a high number of FA remained, which may lead to undesired effects in the removal step. Small dust detections due to noise will be dilated and may lead to an erratic, restored result. The novel detector results in the lowest FAR due to noise, since the co-support operator is applied and effectively lowers these false detections.

The evaluation of the sequence *Surprise* is shown in Figure (5.15). Since this sequence do not contain complex motion, all detectors provide acceptable results. However, the novel dirt and dust detector manage to detect all of the fine struc-

(a) Original image **(b)** GT mask **(c)** SDIa, $\tau = 10$

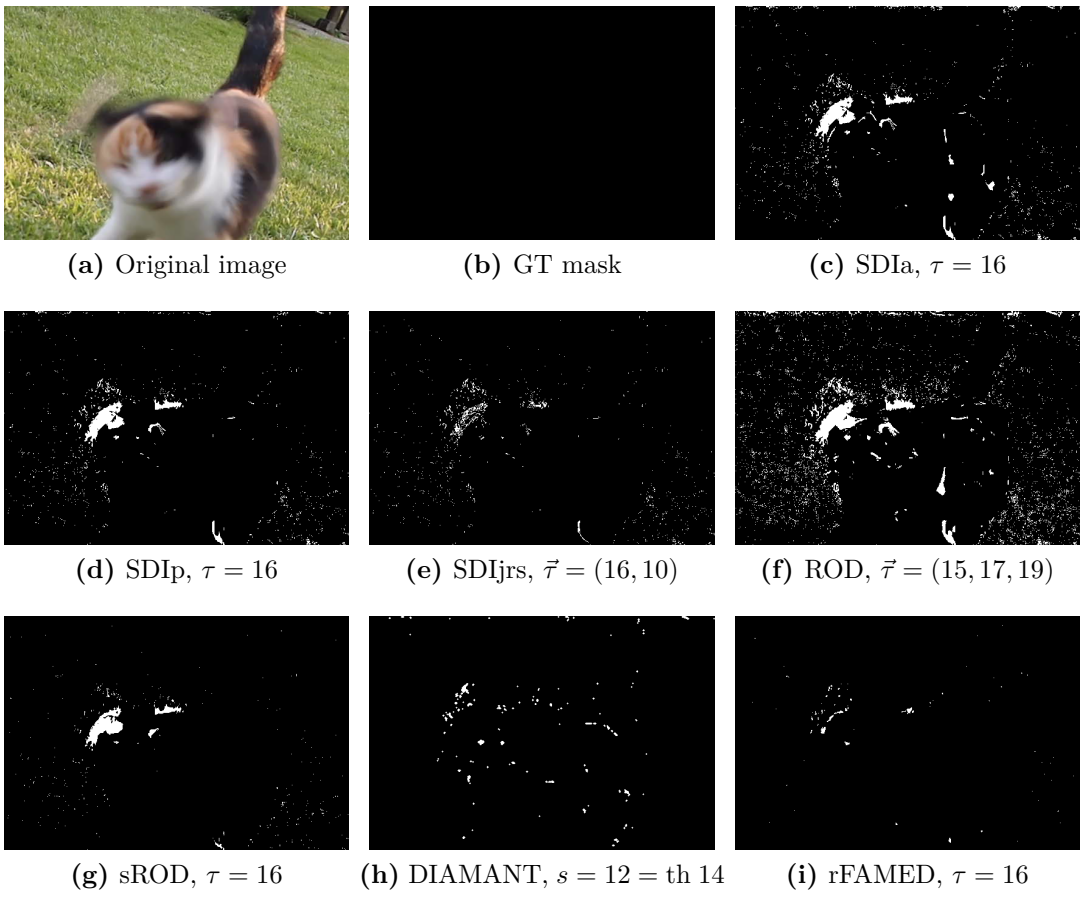
(d) SDIp, $\tau = 10$ **(e)** SDIjrs, $\vec{\tau} = (10, 8)$ **(f)** ROD, $\vec{\tau} = (9, 11, 13)$

(g) sROD, $\tau = 9$ **(h)** DIAMANT, $s = 16$ **(i)** rFAMED, $\tau = 13$

Figure 5.15: Qualitative performance analysis of frame 143 of sequence *Surprise*.

tured hairs, in contrast to all other techniques. Closer inspections of the density damping function may lead to the assumption that larger dust spots are endangered to be damped down as well. This frame proves the contrary, since the large dust spot at the bottom right was detected to its whole extend.

The following sequence contains highly complex motion. Frame 239 of sequence *Cat* and all corresponding dust masks are shown in Figure (5.16). The sequence was shot with a digital camera, therefore no film grain and dust spots are contained.



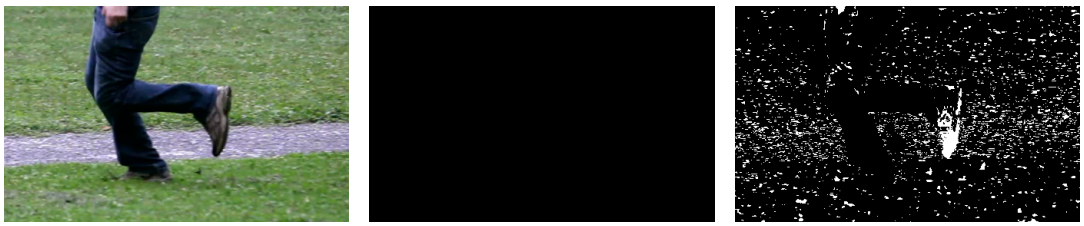
(a) Original image **(b)** GT mask **(c)** SDIa, $\tau = 16$

(d) SDIp, $\tau = 16$ **(e)** SDIjrs, $\vec{\tau} = (16, 10)$ **(f)** ROD, $\vec{\tau} = (15, 17, 19)$

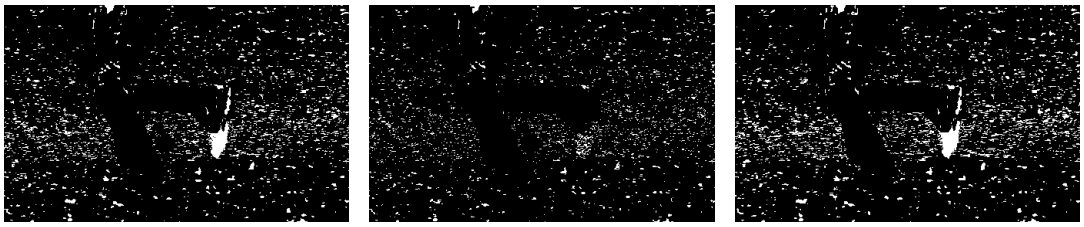
(g) sROD, $\tau = 16$ **(h)** DIAMANT, $s = 12 = \text{th } 14$ **(i)** rFAMED, $\tau = 16$

Figure 5.16: Qualitative performance analysis of frame 239 of sequence *Cat*.

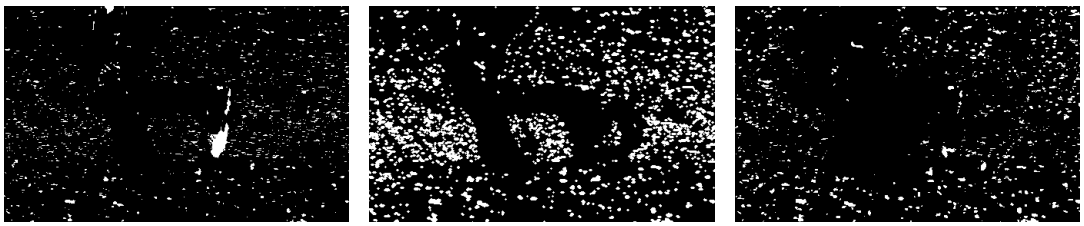
It is a very challenging sequence, not only because it includes highly complex motion, but also due to its highly textured background, which is likely to lead to a hight FAR. When comparing the obtained dust masks, we can conclude, that all temporal and spatio-temporal filtering techniques face severe difficulties. Only the SDIjrs method can avoid a large area of false detection due to its strict condition, that the motion compensated and warped neighboring images have to be similar. The DIAMANT detector results in a lower but still high FAR, especially close to edges. The proposed detector, using the damping functionality in difficult and dangerous areas clearly produce the best result. In the first place, the density measure and its damping technique is essential in such a sequence full of nonrigid



(a) Original image **(b)** GT mask **(c)** SDIa, $\tau = 20$



(d) SDIp, $\tau = 20$ **(e)** SDIjrs, $\vec{\tau} = (20, 12)$ **(f)** ROD, $\vec{\tau} = (19, 21, 23)$

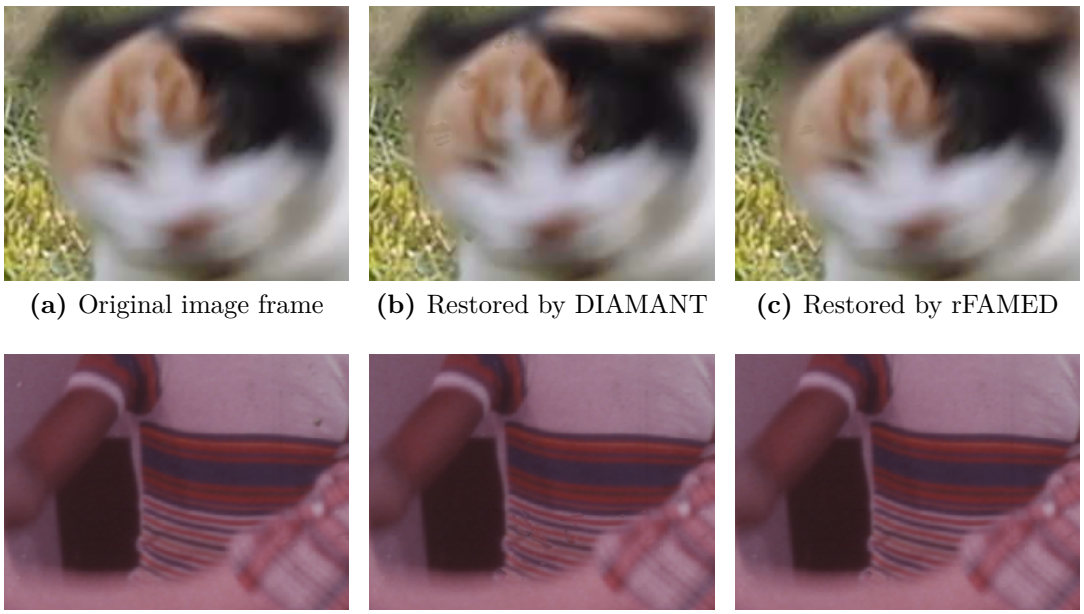


(g) sROD, $\tau = 20$ **(h)** DIAMANT, $s = 8$ **(i)** rFAMED, $\tau = 20$

Figure 5.17: Qualitative performance analysis of frame 32 of sequence *Hannes runs*.

motion. The divergence, as well as the moving edge measure further improve the outcome.

The last example frame, which we want to analyze is chosen from the sequence *Hannes runs* and shown in Figure (5.17). Similar to the previous shown example, this frame also turned out to be challenging due to its difficult background. Each of the examined detector has troubles to discriminate between textured structures and dust spots. However, again, our novel detector results in the best performance, even though false detections still remain in the background. Actually, this sequence contains two kind of motions: Firstly, the local object motion of the running person, and secondly, a translational motion, since the camera is following the moving object. Due to the complicated texture in the moving background, the



(a) Original image frame **(b)** Restored by DIAMANT **(c)** Restored by rFAMED

(d) Original image frame **(e)** Restored by DIAMANT **(f)** Restored by rFAMED

Figure 5.18: Qualitative comparison of restored results. (a) and (d) shows two selected original frames, in (b) and (e) the images were restored by the DIAMANT detector, which results in artifacts, in particular at the head of the cat and in the area of the striped T-shirt, and (c) and (f) show the superior results produced by rFAMED.

motion cannot be estimated correctly. As a consequence, the density measure mainly affects the moving object and mainly damps the responses in this part. By choosing a higher threshold (for DIAMANT a lower sensitivity, respectively) FA in the background area can be reduced but not eliminated completely. The DIAMANT detector even enforces false detections in the background. The SDIjrs detector also yields a relatively low FAR, shown in Figure (5.17e). This observation leads to the conclusion that a strict check for similarity of warped neighboring frames can effectively improve the outcome.

5.4.3 Comparison of restored results

The removal of the dust spots is not part of this diploma thesis, but we have integrated the novel single-frame-defect detector in the DIAMANT restoration

tool in order to show the final result and the superiority of the novel detector over the old technique, denoted as DIAMANT in this diploma thesis. In Figure (5.18), a frame of sequence *Cat* was restored by both approaches DIAMANT and rFAMED and the visually different results are demonstrated.

5.5 Computational Complexity

We have implemented the proposed algorithm in C++. On our machine, an eight-core Xeon processor, 2.67 GHz, 4GB RAM and a NVIDIA GeForce GTX 460 on a 64 bit Windows 7 Professional system, the actual runtime for 2K input frames is on average 780 ms. The DIAMANT detector has an average runtime of 1.57 s for the same input images, which means that the proposed detector is 2 times faster than the previously integrated DIAMANT detector.

5.6 Discussion

In this chapter we have demonstrated the superiority of our novel single-frame-defect detector over all examined techniques. Furthermore, we have achieved the aim, to gain an improvement over the algorithm used in the DIAMANT film restoration software. Mainly, we have succeeded to further remove FA by analysis of PM while preserving fine structured dust spots. Even though, the ROC based evaluation has its limitations due to the used GT, we have given a comparable overview over a variety of detection algorithms, quantitatively as well as qualitatively.

The superiority over other techniques was achieved due to several elements of the novel approach. First we did a primary response measure, which was obtained by a sROD based detection technique. It can be seen in each of the ROC curve, presented in Section (5.4.1), especially in Figures (5.5 - 5.8), that the sROD detector outperforms all other temporal and spatio-temporal filtering methods. Furthermore, in the qualitative performance analysis, which was done in Section (5.4.2), sROD provides a quite good performance. Less single outliers are included and fine structures are better conserved than by any of the temporal filtering methods.

The same principle is followed by the co-support operator, which further improves the results of the sROD based detection. Thus, the response measure of pixels close to potential dust pixels is increased, which mainly results in an expansion of dust spots, and the response values of pixels close to clean pixels are decreased. In other words, this operator controls FA which occur due to noise. The effect of the co-support operator was clearly demonstrated e.g. in Figure (5.15i), where the response of fine structured dust spots has been increased and thus, appears much stronger than in other resulting dust masks. Furthermore, in Figure (5.14i), FA due to noise were clearly reduced when compared with the results of other detectors. The damping functions, such as the density and divergence measure, the moving edge detection measure and the protection measure due to dissimilarity of warped neighbors, bring an essential further improvement. In Figures (5.16i) and (5.17i) the shown results demonstrates the improvements for sequences containing complex motion.

When using challenging image sequences, such as *Dance2*, *Cat* or *Hannes runs* both ME algorithms produce a lot of ME errors and are not able to correctly match and warp the neighboring frames. The resulting effect is a higher FAR in such areas and consequently, there is a higher risk to insert artifacts. The proposed measures damp responses in those areas to a large extent. Remaining FA do not necessarily degrade the quality of the restored result, depending on the size of the false detection and the underlying motion vectors, that are used for the replacement.

For temporal and spatio-temporal filtering methods, we can overall report similar results to those reported in related work. The SDIa method, which was the first motion compensated technique, uses pure pixel differences for the detection. SDIa is able to achieve a high CDR but only at the cost of an increased FAR. The performance further degrades at presence of noise, ME errors or textured regions. The follower, SDIp, performs better, in particular it significantly reduces the number of false detections at presence of noise. Due to its additional constrain of sign consensus between the two warped neighbors it even show an improvement over SDIa close to edges and for small ME errors, which was observed in Figures (5.12) and (5.16). In the qualitative analysis, e.g. in Figure (5.13) or (5.16), it was shown that SDIjrs and sROD produce less FA due to complex motion. An explanation

for that may be, that sROD provides robustness against single outliers, e.g. due to ME errors. The SDIjrs method additionally requires similarity of the two warped neighboring frames. SDIjrs generally is more robust against ME errors, since parts, in which the warped neighbors differ, are treated as unreliable. ROD produces a high amount of FA close to moving edges. The performance of ROD at presence of noise is slightly better than that of SDIp, which can be seen in Figure (5.14), but at presence of textured background, the produced result is even worse than that of SDIa, shown in Figures (5.16f) and (5.17f). The conclusion, that considering more spatial support from temporally neighboring frames leads to a worse performance, was already drawn in [5] and can be confirmed by our results. ROD is said to be generally more robust to ME errors than any of the SDI detectors [10], which we cannot confirm, since the opposite is reflected, e.g. in Figure (5.13). Many FA were caused in textured regions as well as when motion was not accurately estimated. The main limitation of the ROD method is the fact, that it requires the setting of 3 thresholds. sROD in contrast, only uses one threshold and is more robust in textured regions. In recent works, the sROD detection method has been established as basic detection method prior to diverse post processing methods, since it is more robust against noise and small ME errors.

When comparing the results of the novel detector with those of the DIAMANT detector we can observe that the FA due to noise and due to complex motion are significantly lower, which is demonstrated in Figure (5.18), where the restored results by using the computed dust masks of DIAMANT and rFAMED are compared. We have found, that a slight dilation of the dust mask is advantageous, since dust spots usually appear with a smooth transition to the background. Thus, the completeness of the detection depends on the used threshold of the detector. Therefore, the overall performance of these methods can be summarized as follows. The worst ones include SDIa and ROD, followed by SDIp. Then, the others are SDIjrs, sROD, DIAMANT and finally, the best performance is provided by rFAMED.

An interesting aspect is that surprisingly, the detector's performance obviously is not dependent on the used ME algorithms. Even though we slightly adapted the parameterization for the Pixel motion, in order to apply the damping in a stricter way, the curves are overall similar for every sequence.

In addition, the runtime is satisfying since we avoid complex calculations and the costly motion compensation was accelerated on a GPU. All other image operations were implemented efficiently as well.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this master's thesis a novel approach for single-frame-defect detection in archived film material is proposed. The aim was to develop a new state of the art method, which particularly focuses on minimizing false detections since falsely detected regions might introduce artifacts into the restored result to contain artifacts. At the same time, the correct detection rate has been kept clearly in mind.

Generally, dirt and dust detection appears to be difficult in areas of complex motion or in areas of motion discontinuities. To gain a satisfying result for all kind of archived film material, ME is an essential component for the detection method. Although, many FA may occur due to ME errors, we use motion compensated neighboring frames, since otherwise, the consideration of complex motion, occlusion and uncovering areas of locally moving objects is simply impossible. Four ME algorithms have been used within this work and two of them were further examined within the evaluation of varying detection methods. Experiments have shown that the dependency between the chosen ME algorithm and the obtained results is less than expected, since overall a similar performance has been demonstrated for the different ME approaches.

The novelty of our approach lies in the consideration of PM when detecting single-frame-defects. We present a combined solution, consisting of three independent steps. First, a primary detection method based on the sROD detector

is integrated, which results in a probability distribution indicating the likelihood for each pixel of being defected. This information about each single pixel is kept and utilized over the whole detection process. The sROD method effectively reduces single outliers and at the same time, it is robust against noise and small ME errors. Furthermore, we introduce the co-support operator, which is an effective tool to increase the robustness against noise and film grain. The dilatation effect completes dust areas and further improves the results. Finally, we have applied several measures indicating areas that are likely to produce FA. Those measures are inferred by analyzing the motion vector fields and the warped neighboring frames. As a result of deductions made from the motion field, areas of PM are obtained and the response measure is damped within these areas. Consequently, for all potential dust spots found in these areas, the probability of being defected is decreased, dependent on the strength of the applied measure. Several measures are incorporated, which have been configured in order to best fit the characteristics of the chosen ME algorithm.

The rFAMED technique has been tested on real corrupted image sequences and experiments have shown an overall better performance for rFAMED, than for all other examined techniques including the DIAMANT detector. The benefit of the proposed detection method is twofold: First, rFAMED provides the ability to detect fine structured defects without involving an exploding number of FA and secondly, areas of PM are correctly found, which leads to a significant reduction of annoying artifacts in the restored results. The superiority against other techniques has been demonstrated in terms of ROC performance and visually as well. Furthermore, rFAMED convinces by its low computational complexity.

6.2 Future work

We have presented several measures in order to detect areas of PM, however only a subset was chosen for the final algorithm. Some measures, such as the measure of tracing of occluded/uncovered motion vectors as well as the magnitude measure have potential to be further examined. The magnitude measure could be extended to a discrimination between local and global motion, in order to deal with both separately.

The majority of the proposed damping measures, in particular those indicating occlusion areas, behave similar around large-sized blotches (e.g. 5% of image size) as well as in areas of occlusions and uncoverings. Since those response values are damped down, large defected areas are not detected. This is a limitation of the novel detection technique, which has to be further investigated in future work.

Moreover, there are some motivations to develop further damping methods. For example, a measure can be incorporated, that compares intensity values of center and the target position in the appropriate neighbor by following the motion vector. This will indicate dust spots in the neighboring frame and could be utilized for the detection of large blotches as well. Furthermore, combined measures could provide valuable information, such as the magnitude measure combined with an edge detector: Motion vectors with a high magnitude indicate moving parts and high gradients within those areas indicate textured regions. The joint occurrence of both is likely to result in false detections.

Usually, only parts within a frame contain complex, local object motion. Thus, a further refinement regarding computational complexity is suggested. By inferring areas of local and complex motion from the motion vector field, the whole damping approach and following complicated and time-consuming post processing steps could be limited to those parts only.

Since the whole approach consists of widely independent parts, every component could be exchanged or additional techniques could be applied, such as post processing methods proposed in related work, which are available at an immense diversity.

Finally, there is potential to develop a new evaluation schema specialized for the dirt and dust detection domain. As already suggested in [5] an evaluation schema based on the comparison of intensity values rather than of binary masks might lead to a more reflective representation of achieved performances. Furthermore a blotch based evaluation schema with an integrated weighting function, based on the blotch-size, can be introduces. However, the problem of line scratches in the GT remains, but the difficulty of choosing a threshold for the IR scan and the deviation due to missed and falsely detected defects resulting from the incompleteness of the GT are resolved.

List of Acronyms

AR auto-regressive

CD correct detections

CDR correct detection rate

CPU central processing unit

DFD displaced frame difference

DTMF double-threshold median filter

FA false alarms

FAR false alarm rate

FOLKI Iterative Lucas-Kanade optical flow

GN Gauss-Newton

GPU graphics processing unit

GT ground truth

HFVM Hierarchical Feature Vector Matching

HOT histogram of template

IR infrared

LK Lucas-Kanade

LUM Lower–Upper–Middle

LUM' 3-frame-Lower–Upper–Middle

MAP maximum a posteriori

ME motion estimation

ML3D multi-level median filter in 3 sub-windows

ML3Dex multi-level median filter in 5 sub-windows

MLF multi-level median filter

MOS multi-stage order statistic

MRF Markov random field

PM pathological motion

rFAMED single-frame-defect detection approach based on the Reduction of FA by analysing the estimated Motion followed by an Extraction of Damping functions

ROC Receiver operating characteristics

ROD rank-ordered difference

SDIa Spike detection index a

SDIjrs Spike detection index (version developed by JOANNEUM RESEARCH)

SDIp Spike detection index p

SMF soft morphological filtering

sROD simplified rank-ordered difference

SSMF standard spatial median filter

TV total variation

Bibliography

[1] JOANNEUM RESEARCH, "Diamant - film restoration." `http://diamant.joanneum.at/film_restoration`, Oct. 2011.

[2] Filmarchiv Austria. `http://filmarchiv.at`, Oct. 2011.

[3] A. Kuiper and M. Sigmund, "Simulating of Authentic Movie Faults," *The International Conference on Computer as a Tool, EUROCON 2005*, vol. 2, pp. 1015–1018, 2005.

[4] A. Buadés, J. Delon, Y. Gousseau, and S. Masnou, "Adaptive Blotches Detection for film restoration," *17th International Conference on Image Processing (ICIP)*, pp. 3317 – 3320, 2010.

[5] J. Ren and T. Vlachos, "Detection of dirt impairments from archived film sequences: survey and evaluations," *SPIE Optical Engineering*, vol. 49, June 2010.

[6] P. van Roosmalen, J. Biemond, and R. Lagendijk, "Restoration and storage of film and video archive material," *Signal Processing for Multimedia*, 1999.

[7] R. L. Lagendijk, P. M. B. van Roosmalen, and J. Biemond, "Video enhancement and restoration," in *Handbook of Image and Video Processing*, pp. 227–241, Academic Press, 2000.

[8] A. Licsar, L. Czuni, and T. Sziranyi, "Trainable Postprocessing Method to Reduce False Alarms in the Detection of Small Blotches of Archive Films," *International Conference of Image Processing (ICIP)*, pp. 562–565, 2005.

[9] M. A. Ahmed, F. Pitié, and A. C. Kokaram, "Extraction of non-binary blotch mattes," *International Conference of Image Processing (ICIP)*, pp. 2757 – 2760, 2009.

[10] A. C. Kokaram, "On missing data treatment for degraded video and film Archives: a survey and a new Bayesian approach," *IEEE Transactions on Image Processing*, vol. 13, no. 3, p. 397–415, 2004.

[11] A. Nieminen, P. Heinonen, and Y. Neuvo, "A new class of detail-preserving filters for image processing," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. PAMI-9, no. 1, pp. 74–90, 1987.

[12] J. Ren and T. Vlachos, "Dirt detection for archive film restoration using an adaptive spatio-temporal approach," *The 2nd IEE European Conference on Visual Media Production (CVMP)*, pp. 219–228, 2005.

[13] Buisson O. and Besserer B. and Boukir S., "Deterioration Detection for digital film restoration," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 78–84, 1997.

[14] Tilie S. and Bloch I. and Laborelli L., "Fusion of complementary detectors for improving blotch detection in digitized films," *Pattern Recognition Letters*, vol. 28, no. 13, pp. 1735–1746, 2007.

[15] J. Gallagher, N. and G. Wise, "A theoretical analysis of the properties of median filters the human interaction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 6, pp. 1136 – 1141, 1981.

[16] R. Hardie and C. Boncelet, "LUM Filters: a class of rank-order-based filters for smoothing and sharpening," *IEEE Transactions on Signal Processing*, vol. 41, no. 3, p. 1061–1076, 1993.

[17] S. Tilie, L. Laborelli, and I. Bloch, "A contrario False Alarms Removal for Improving Blotch Detection in Digitized Films Restoration," *6th Conference on Signals and Image Processing (EURASIP)*, pp. 410–413, 2007.

[18] R. Storey, "Electronic detection and concealment of film dirt," *Journal of Society of Motion Picture and Television Engineers*, vol. 94, pp. 642–647, 1985.

[19] Gangal A. and Kayikçioglu T. and Dizdaroglu B., "An improved motion-compensated restoration method for damaged color motion picture films," *Signal Processing Image Communication*, vol. 19, no. 4, p. 353–368, 2004.

[20] A. C. Kokaram, *Motion Picture Restoration*. Springer-Verlag (Berlin), 1998.

[21] P. Schallauer, A. Pinz, and W. Haas, "Automatic restoration algorithms for 35 mm film," *Journal of Computer Vision Research*, vol. 1, no. 3, pp. 59–85, 1999.

[22] P. Schallauer, "Digital Image Sequence Restoration," Master's thesis, University of Technology, Graz, 1996.

[23] Arce G. R., "Multistage order statistic filters for image sequence processing," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, p. 1146–1163, 1991.

[24] B. Alp, P. Haavisto, T. Jarske, K. Oistamo, and Y. Neuvo, "Median-based algorithms for image sequence processing," *SPIE Visual Communications and Image Processing*, pp. 122–134, 1990.

[25] J. Ren and T. Vlachos, "Efficient detection of temporally impulsive dirt impairments in archived films," *Signal Processing*, vol. 87, no. 3, p. 541–551, 2007.

[26] M. S. Hamid, N. R. Harvey, and S. Marshall, "Genetic algorithm optimization of multidimensional grayscale soft morphological filters with applications in film archive restoration," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 13, pp. 406 – 416, 2003.

[27] Nadenau M. J. and Mitra S. K., "Blotch and scratch detection in image sequences based on rank ordered differences," *Proceedings of the 5th International Workshop on Time-Varying Image Processing and Moving Object Recognition*, p. 27–35, 1996.

[28] S. Wei, R. Zhang, P. Hao, and Y. Ding, "Blotch Detection Based on Texture Matching and Adaptive Multi-Threshold," *5th International Conference on Image and Graphics*, 2009.

[29] M. K. Güllü, O. Urhan, and S. Ertürk, "Blotch detection and removal for archive film restoration," *International Journal of Electronics and Communications (AEU)*, p. 534 – 543, 2008.

[30] P. M. B. van Roosmalen, "Restoration of Archived Film and Video," tech. rep., Technical University Delft, 1999.

[31] J. Ren and T. Vlachos, "Non-motion-compensated region based dirt detection for film archive restoration," *Optical Engineering*, vol. 45, no. 8, 087004, 2006.

[32] J. P. Biemond, P. van Roosmalen, and R. L. Lagendijik, "Improved blotch detection by postprocessing," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 6, p. 3101–3104, 1999.

[33] A. C. Kokaram and S. J. Godsill, "MCMC for joint noise reduction and missing data treatment in degraded video," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, p. 189–205, 2002.

[34] A. Kokaram, "Practical MCMC for missing data treatment in degraded video," *in Proceedings of Eur. Conf. Computer Vision Workshop Statistical Methods Video Processing*, p. 85–90, 2002.

[35] S. Kalra, M. Chong, and D. Krishnan, "A New Auto-Regressive (AR) Model-Based Algorithm for Motion Picture Restoration," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, p. 2557–2560, 1997.

[36] M. N. Chong and D. Krishnan, "An edge-preserving MRF model for the detection of missing data in image sequences," *Signal Processing Letters*, vol. 5, no. 4, pp. 81–83, 1998.

[37] A. C. Kokaram, R. Morris, W. J. Fitzgerald, and P. J. W. Rayner, "Detection of missing data in image sequences," *IEEE Transactions on Image Processing*, vol. 4, no. 11, p. 1496–1508, 1995.

[38] S.-C. Nam, M. Abe, and M. Kawamata, "Fast and efficient MRF-based detection algorithm of missing data in degraded image sequences," *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, vol. E91-A, no. 8, pp. 1898–1906, 2008.

[39] M. Ghaderi and S. Kasaei, "Novel Post-Processing Methods Used in Detection of Blotches in Image Sequences," *International Journal of Electronics and Communications (AEU)*, vol. 58, pp. 58–64, 2004.

[40] Z. Xu, H. R. Wu, and X. Yu, "Object Features Based Spatial and Temporal Blotch Detection for Archive Video Restoration," *IEEE Transactions on Image Processing*, 2010.

[41] A. Licsár and L. C. Tamás Szirányi, "Trainable blotch detection on high resolution archive films minimizing the human interaction," *Machine Vision and Applications*, vol. 21, no. 5, pp. 767–777, 2009.

[42] Z. Xiaona, Q. Guoqing, X. Rong, and Z. Tao, "An Improved Approach of Detection and Restoration Blotches in Archived Films," *International Congress on Image and Signal Processing (CISP)*, pp. 1–5, 2009.

[43] Rares A., Reinders J. T., and Biemond J., "Statistical Analysis of Pathological Motion Areas," *IEEE Seminar on Digital Restoration of Film and Video Archives*, 2001.

[44] D. Corrigan, N. Harte, and A. Kokaram, "Pathological Motion Detection for Robust Missing Data Treatment in Degraded Archived Media," *International Conference of Image Processing (ICIP)*, pp. 621–624, 2006.

[45] B. Kent, A. Kokaram, B. Collis, and S. Robinson, "Two layer segmentation for handling pathological motion in degraded post production media," *International Conference of Image Processing (ICIP)*, vol. 1, pp. 299–302, 2004.

[46] A. Rares, J. T. Reinders, and J. Biemond., "Complex event classification degraded image sequences," *International Conference of Image Processing (ICIP)*, October 2001.

[47] L. Alvarez, R. Deriche, T. heo Papadopoulo, and J. S. anchez, "Symmetrical Dense Optical Flow Estimation with Occlusions Detection," *International Journal of Computer Vision*, vol. 75, no. 3, p. 371–385, 2007.

[48] J. Sun, Y. Li, and S. B. Kang, "Symmetric Stereo Matching for Occlusion Handling," *omputer Vision and Pattern Recognition*, pp. 399–406, 2005.

[49] D. Feldman and D. Weinshall, "Motion Segmentation and Depth Ordering Using an Occlusion Detector," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 30, pp. 1171–1185, July 2008.

[50] K. P. Lim, A. Das, and M. N. Chong, "Estimation of occlusion and dense motion fields in a bidirectional Bayesian framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 5, pp. 712–718, 2002.

[51] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1293 –1300, june 2010.

[52] J. Ren and T. Vlachos, "Segmentation-assisted detection of dirt impairments in archived film sequences," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 2, p. 463–470, 2007.

[53] C. Rhemann, M. Bleyer, and M. Gelautz, "A Graph-Based Approach to Optical Flow Estimation," *Junior Scientist Conference, Vienna, Austria*, pp. 61–63, 2006.

[54] C. Stiller and J. Konrad, "Estimating motion in image sequences," *Signal Processing Magazine, IEEE*, vol. 16, pp. 70–91, july 1999.

[55] D. Corrigan, "Motion Estimation Reliability and the Restoration of Degraded Archived Film," *PhD-Thesis*, 2007.

[56] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, no. 17, pp. 185–203, 1981.

[57] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Annual Symposium of the German Association for Pattern Recognition (DAGM)*, vol. 4713, pp. 214–223, 2007.

[58] J. Xiao, H. Cheng, H. Sawhney, C. Rao, M. Isnardi, and S. Corporation, "Bilateral filtering-based optical flow estimation with occlusion detection," in *European Conference on Computer Vision (ECCV)*, vol. 1, pp. 211–224, 2006.

[59] S. Baker, D. Scharstein, and J. Lewis, "A Database and Evaluation Methodology for Optical Flow," *in Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2011.

[60] D. Fleet and Y. Weiss, "Optical Flow Estimation," *in Mathematical models for Computer Vision: The Handbook*, 2006.

[61] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, "Anisotropic Huber-L1 Optical Flow," *in Proceedings of British Machine Vision Conference (BMVC)*, 2009.

[62] G. Paar and W. Pölzleitner, "Robust disparity estimation in terrain modeling for spacecraft navigation," *11th International Conference on Pattern Recognition (IAPR)*, pp. 738 – 741, 1992.

[63] G. L. Besnerais and F. Champagnat, "Dense optical flow by iterative local window registration," *IEEE International Conference on Image Processing*, 2005.

[64] F. Champagnat, A. Plyer, G. L. Besnerais, B. Leclaire, S. Davoust, and Y. L. Sant, "Fast and accurate PIV computation using highly parallel iterative correlation maximization," *Experiments in Fluids*, pp. 1169–1182, 2011.

[65] S. Tang and S. Goto, "Histogram of template for human detection," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2186 – 2189, 2010.

[66] C. Bartels and G. De Haan, "Occlusion Classifiers for Picture Rate Conversion," *Proceedings of SPIE*, pp. 72571D–72571D–8, 2009.

[67] C. D. Brown and H. T. Davis, "Receiver operating characteristics curves and related decision measures: A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 80, no. 1, pp. 24–38, 2006.