Segmentation-Assisted Detection of Dirt Impairments in Archived Film Sequences

Jinchang Ren and Theodore Vlachos

Abstract—**A novel segmentation-assisted method for film dirt detection is proposed. We exploit the fact that film dirt manifests in the spatial domain as a cluster of connected pixels whose intensity differs substantially from that of its neighborhood and we employ a segmentation-based approach to identify this type of structure. A key feature of our approach is the computation of a measure of confidence attached to detected dirt regions which can be utilized for performance fine tuning. Another important feature of our algorithm is the avoidance of the computational complexity associated with motion estimation. Our experimental framework benefits from the availability of manually derived as well as objective ground truth data obtained using infrared scanning. Our results demonstrate that the proposed method compares favorably with standard spatial, temporal and multistage median filtering approaches and provides efficient and robust detection for a wide variety of test material.**

Index Terms—**Film dirt detection, archive restoration, image segmentation.**

I. INTRODUCTION AND BACKGROUND

DIRT or sparkle is among the most commonly encountered impairments in archived film material and consequently its successful detection is a priority issue in any archive restoration system [1-2], [4]. Dirt may be airborne during film processing, may be due to fibers from clothing or human tissue, may be due to wear and tear of moving mechanical parts or even due to fragments of the film itself owing to excessive friction as the film moves at speed in contact with mechanical components. Dirt particles can adhere to film at any processing stage like exposure, development, printing, telecine transfer or any of the intervening winding and handling operations. Dirt manifests itself in one of two ways; a dirt particle adhering to negative film will eventually appear as bright when printed or reproduced as a positive image while a dirt particle adhering to positive film will obviously impede the transmission of light and will appear dark.

J. Ren is on leave from the School of Computers, Northwestern Polytechnic University, Xi'an, 710072, China. He is now with the Centre for Vision, Speech and Signal processing (CVSSP), University of Surrey, U.K. (email: j.ren@surrey.ac.uk, npurjc@nwpu.edu.cn).

T. Vlachos is with CVSSP, University of Surrey, U.K. (Corresponding author, phone: +44-1483-689854, fax: +44-1483-686031, e-mail: t.vlachos@surrey.ac.uk).

Conventional (non-computer-assisted) detection of dirt impairments is carried out by optical processing which requires a specially designed telecine and relies on the transparency of film emulsions to infrared light allowing dirt particles to be detectable by an infrared sensor. This has a number of practical limitations including the fact that detection is not very effective for particles of small size due to limitations of lens performance at infrared wavelengths. This technique is applicable exclusively to colour film because the silver image in monochrome film is opaque to infrared radiation.

General speaking, dirt is a temporally impulsive (single-frame) event, appearing mostly as dark or bright opaque spots of random size, shape and location (Fig. 1). In non-pathological cases a dirt particle of a given shape and size will normally be present at a specific spatial location of a particular frame. Due to the above interframe processing has proved a useful tool towards detection and concealment [2], [5].

(a) Static text   (b) Low-motion natural scene

(c) Moderate-motion natural scene   (d) Fast-motion natural scene
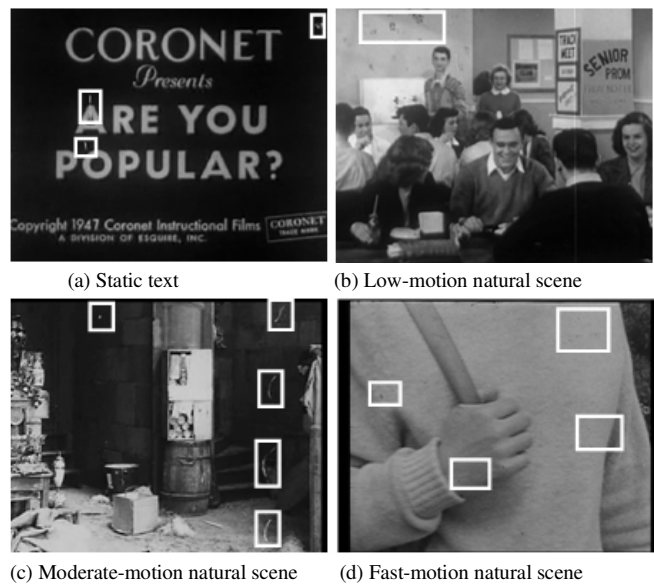
Fig. 1. Examples of dirt (marked in white boxes).

Since motion-compensated prediction requires a high degree of complexity and can be unreliable when motion estimation fails, many spatial filtering techniques have also been proposed as alternatives. Existing methods and models on dirt detection with or without motion compensation are discussed below.

A. Dirt Detection without Motion Compensation

Storey's work [15] was perhaps the earliest reported contribution targeting specifically the electronic detection and concealment of film dirt. According to this approach a pixel is flagged as dirt if the corresponding absolute differences between the current frame and each of the previous and next frames are high. Other non-motion-compensated methods employ median or morphological filtering [7, 9-12].

Nieminen *et al* [9] presented a multi-stage median filter (MMF) which uses hierarchical median operations to reject sparkle type of distortions. In Arce [10], MMF filters have further evolved as multi-stage order statistic filters (MOS). Senel *et al* [11] proposed a topological median filter to extract edges in noise; however, the filtered images are of unacceptable

visual quality in most cases. Hardie, and Boncelet [12], proposed LUM (lower-upper-middle) filters, which utilized two parameters for adjustable smoothing and sharpening of images, but ultimately proved problematic towards the restoration of scenes containing fast-moving objects [7]. Buisson *et al* [28] considered morphological tools but eventually had to rely on motion estimation in their combined solution.

Most of the methods in this category are based on pixel-level or small-window-level operations, which means that candidate dirt pixels are located in a relatively small neighborhood. In contrast, work by Kokaram *et al* [21] demonstrated a capability of coarse detection of large areas of defect provided that there are noticeable changes in local histograms. In both median and morphological filtering approaches, an unsuitable size of window may cause many false alarms with heavy degradation of visually significant edges [29]. To solve this problem, Hamid *et al* proposed an alternative based on soft morphological filtering (SMF) and used genetic algorithms to determine the size and shape in the filters [7]. Nevertheless, SMF seems impractical for most applications because it needs a considerable amount of representative dirt samples for training purposes before filtering parameters can be optimized.

B. Dirt Detection with Motion Compensation

A logical extension to Storey's approach would be the use of bi-directional motion compensation commonly complemented by temporal median filtering using the current frame and its two motion-compensated neighbors [2]-[5]. More complex methods require the availability of longer motion trajectories [14]. The concept of motion-compensated temporal filtering comes as early in 1980's [30]. In [31], Dubois and Sabri applied this concept to noise reduction in television signals. Their work was further extended to include the use of recursive filters [32,33].

In Kokaram's work [4], the "Spike Detection Index" (SDI) identifies high-valued absolute differences between the current and two motion compensated images. The extended SDI method, SDIp, additionally requires sign consensus from the two differences above. In Schallauer *et al* [2], it is examined whether both the absolute differences between a pixel and co-sited pixels in the two motion-compensated neighboring frames exceed a first (higher) threshold while at the same time the absolute difference between the two motion-compensated frames is less than a second (lower) threshold. Nadenau, and Mitra [5], have proposed the rank order detector (ROD), in which a total of seven pixels from three consecutive frames are compared against three thresholds. Gangal *et al* extended ROD to five frames to improve accuracy in heavily corrupted images or occluded blotches [3].

In the framework of motion compensation, model-based approaches can be used, such as Wiener filtering, AR (auto-regressive), MRF (Markov random filed), Gibbs distribution, and Gibbs-Markov random fields [3], [8], [16], [18], [22-27]. The determination of a MRF prior allows the detection of dirt in a Bayesian framework [1], [19], [20]. Since definitive statistical models are difficult to obtain, all the above 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 [23], [24].

C. Main Contributions

In this paper, we propose a segmentation-assisted method for the detection of dirt in archived films. Relative to other methods in the literature, a key feature of our proposal is segmentation allowing the detecton of dirt at a higher semantic level as a *region* (consisting of connected pixels) and not as a loose collection of isolated pixels. This is consistent with the actual manifestation of dirt in real film samples. A second attractive feature of our scheme is that a confidence measure is derived and attached to detected dirt regions. This is an invaluable feature towards both automatic and operator-assisted dirt concealment as it allows performance fine tuning according to preference especially towards achieving a desirable balance between false alarm and correct detection rates. Finally, our method does not employ motion estimation and motion compensated prediction which reduces its complexity considerably and makes it a good candidate for fast implementations.

This paper is organized as follows. In Section II, we introduce our segmentation-assisted approach while Section III contains the definition of a confidence measure and its use towards fine tuning detection performance. Section IV provides experimental evidence arising from testing the proposed scheme on archived film sequences while in Section V the influence of parameter selection is discussed. In Section VI, we draw brief conclusions while detection details using spatial, temporal and multistage median filtering are presented as an Appendix.

II. DIRT DETECTION USING SEGMENTATION

In spatial terms, dirt manifests as a cluster of connected pixels whose intensity differs, on the average, from that of its immediate neighbourhood. Segmentation is a useful tool towards the identification of such a structure because the latter is expected to emerge as an island region in a sea of neighbouring regions corresponding to dirt-free picture content. Our work is one of the few that have appeared in the literature adopting a segmentation methodology for dirt detection, and other work can be identified involving motion field segmentation [17] and watershed methodology [22] in such a context. For simplicity, throughout this work we will consider the luminance (Y) component of images.

A. Image Segmentation by Seedless Region-Growing

We have employed seedless region growing which, contrary to conventional seeded region growing [13], does not require the specification of seed points. Instead, we use a raster scan order to identify any previously unmarked pixel. Such a pixel is then used as a seed for region growing. All pixels previously merged to form a region are marked so they are not revisited.

Region growing is controlled by examining adjacent pixel similarity. An unmarked pixel p which is adjacent to a pixel q of that region will be merged in that region if its intensity $f(p)$ satisfies both the following conditions,

$$f(p) \in [\mu - 3\delta, \mu + 3\delta] \qquad (1)$$

$$f(p) \in [f(q) - 0.8\delta, f(q) + 0.8\delta] \qquad (2)$$

where μ is the sample mean of that region, δ is an acceptance threshold enforcing similarity and $f(q)$ is the intensity value of pixel q. Eq. (1) requires that p has an intensity value close to the mean intensity of the whole region while Eq. (2) enforces close similarity between $f(p)$ and its potential neighbors.

Whenever a new pixel is merged to a region, μ of that region is updated. Parameter δ is determined empirically as $\delta = \sqrt{\mu} / \lambda$ to enable wider acceptance with increasing μ. Parameter λ controls segmentation granularity and useful values for it were experimentally determined in the range $2 \le \lambda \le 6$. As λ increases δ will decrease, more regions will emerge and consequently more detail will be preserved in the segmented image. Finally, the average intensity, i.e. μ, is assigned to all the pixels within the segmented region.

Fig 2 shows two segmentation examples obtained from Fig 1(b) for two different values of λ. For comparison, enlarged part of the original image and the difference between two segmented results are also given. We can see that larger λ leads

(a) $\lambda = 2$ (b) $\lambda = 4$
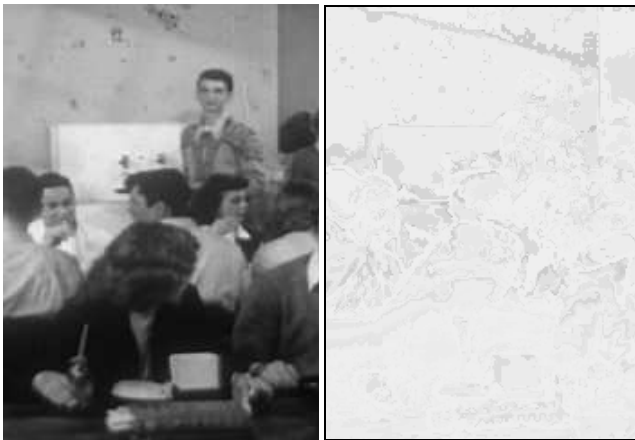

(c) Enlarged original image   (d) Difference between (a) and (b)

Fig. 2. Segmentation of Fig 1(b) for different values of λ.

to finer segmentation in which regions of dirt pixels form clusters as expected. On the other hand, this may occasionally cause higher false alarm rates as discussed in Section V.

B. Determination of Candidate Dirt Locations

We assume that region size is a useful criterion towards an initial identification of candidate dirt locations. For this purpose we use two thresholds operating on region size: a lower one whose purpose is to filter out tiny regions likely to correspond to noise and an upper one s whose purpose is to filter out sizeable regions likely to correspond to actual scene objects (or parts thereof). Again it should be noted that our experimental evidence suggests that segmentations obtained from sharpened images are less sensitive to the choice of such threshold values. Fig. 3 illustrates the effect of different upper thresholds on initial detection. Detected regions (collectively referred to as a detection mask) are shown in white and were derived using the original image shown in Fig 1(b). These results were obtained using two different upper threshold values s and a lower threshold value of 6. We can see that there are still many false alarms in the detection mask, mainly due to moving edges or static small objects. In the next section, interframe information (non-motion-compensated) will be employed to improve the accuracy of the detection mask.
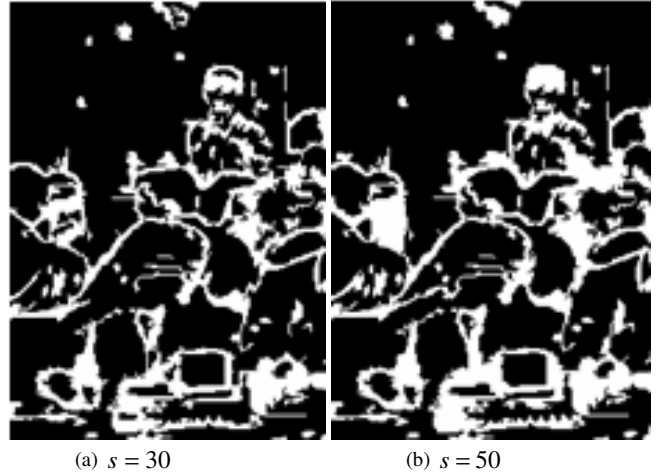
(a) $s = 30$ (b) $s = 50$

Fig. 3. Initial dirt masks detected from Fig 2(b) with size threshold $s = 30$ pixels (a) and $s = 50$ pixels (b).

III. CONFIDENCE-BASED WEIGHTING

A. Extracting Confidence Measurement

The fundamental assumption that dirt is a single-frame event leads naturally to the idea of using inter-frame information for validation. Let f_{n-1}, f_n and f_{n+1} be three consecutive frames. We define $d_{n-} = f_n - f_{n-1}$ and $d_{n+} = f_n - f_{n+1}$ as the forward and backward frame differences, respectively. We also consider the absolute values of d_{n-} and d_{n+} as fundamental confidence indicators. We define d_n as:

$$d_n = \begin{cases} \dfrac{2d_{n-}d_{n+}}{|d_{n-}|+|d_{n+}|} & if \quad d_{n-}d_{n+} > 0 \\ 0 & otherwise \end{cases} \qquad (3)$$

This attains its maximum value when an idealized dirt impulse occurs against a constant background i.e. when $d_{n-} = d_{n+}$. If both d_{n-} and d_{n+} are negative or positive, this relates respectively to dark or bright dirt pixels (particles adhered on negative or positive film stock).

Furthermore, to each value m in d_n ($m \in [0, L-1]$, where L is the total number of grey levels), an associated probability can be defined as follows:

$$p_n(m) = \begin{cases} \lambda_0^{-1} \sum_{k=m_0}^{m} p_d(k) & if \quad k \geq m_0 \\ 0 & otherwise \end{cases} \qquad (4)$$

$$\lambda_0 = 1 - \sum_{k=0}^{m_0-1} p_d(k) = \sum_{k=m_0}^{L-1} p_d(k) \qquad (5)$$

In (4) and (5), p_d is the sample probability density function (PDF) of d_n, (i.e the histogram of d_n). Parameter λ_0 is used to normalize $p_n(m)$ within [0,1] and m_0 to control the removal of static background.

Let μ_0, γ and σ_0 be the sample mean, median and variance of the distribution of values in d_n and let m_0 be defined as a weighted combination of these parameters

$$m_0 = w_\mu \mu_0 + w_\gamma \gamma + w_\sigma \sigma_0 \qquad (6)$$

where $w_\mu, w_\gamma, w_\sigma > 0$. The set of values $w_\mu = w_\gamma = 1/2$, and $w_\sigma = 1$ yielded particularly good results in our experiments.

A graphic interpretation of the above parameters for a sample one-sided, monotonic distribution (typically a reasonable approximation for p_d) is shown in Fig. 4. It is worth noting that
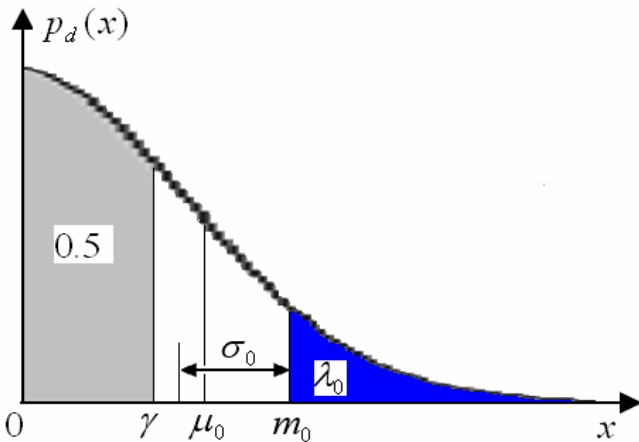

Fig. 4. Relationships among the parameters in Eq. (4) to (6).

setting $m_0 = 0$ (i.e. $\lambda_0 = 1$) in Eq. (4) amounts to histogram equalization of d_n. On the other hand, a static background in three consecutive frames may force most pixel values in d_n near zero, which may lead the incorrect assignment of a high confidence value to a low valued m. Therefore, straightforward histogram equalization is not useful in this context.

Using p_n a direct confidence measurement is obtained as $Conf(i,j) = p_n(d_n(i,j))$, this can be further modified in Eq. (7) for enhanced visibility of candidate dirt pixels.

$$Conf_n(i,j) = \ln(1 + Conf(i,j))/\ln 2 \qquad (7)$$

Fig. 5(a) gives an example of a confidence image obtained for Fig. 1(b). We can see that dirt pixels are identified with high confidence (bright greylevels), though there remain some false alarms mainly due to motion and moving edges.

(a) Confidence image for Fig 1(b)   (b) Binary dirt mask for Fig 1(b)

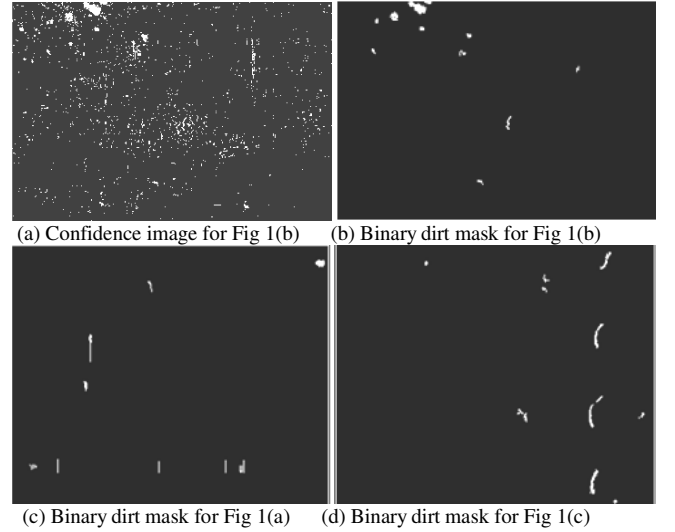(c) Binary dirt mask for Fig 1(a)   (d) Binary dirt mask for Fig 1(c)

Fig. 5. Confidence image for Fig 1(b) with detected binary mask of dirt from Fig 3(b) after confidence-based weighting with $\eta = 2.5$. Masks of dirt detected from Fig 1(a) and Fig. 1(c) are also shown in (c) and (d).

B. Confidence-based Weighting

To improve detection accuracy we combine B_m, the binary dirt mask obtained in Section II, and $Conf$, the confidence image extracted above as follows.

Firstly, a new image $Conf_d$ is obtained as

$$Conf_d(i,j) = \begin{cases} Conf(i,j) & if \quad B_m(i,j) \neq 0 \\ 0 & otherwise \end{cases} \qquad (8)$$

The above rule simply enforces zero confidence for all undetected (i.e. not belonging to the mask B_m) dirt regions.

Secondly, we attempt to remove false alarms due to the movement of small objects. An efficient way to achieve this is to obtain two binary masks, B_c and B_d, by thresholding $Conf$ and $Conf_d$ using a relatively high-valued threshold (i.e. 85% of

the available grey-scale). Subsequently pairs of corresponding regions in the two binary masks are considered. For each connected component of size s_d in B_d, any non-zero pixel in this region is used as a seed to try and locate another connected component of size s_c in B_c. From Eq. (8), we have $Conf_d \subset Conf$, thus we always have $B_d \subset B_c$ and $s_d \le s_c$. For a given parameter $\eta \in [1.5,3]$, the region in $Conf_d$ is assumed to have been the result of a false alarm if condition $s_d * \eta \le s_c$ is satisfied.

Finally, a binary dirt detection mask D_q can be determined under a given confidence level c_q as

$$D_q(i,j,c_q) = \begin{cases} 1 & if \quad Conf_d(i,j) \ge c_q \\ 0 & otherwise \end{cases} \quad (9)$$

In general, c_q can be used to control the sensitivity of detection. One binary mask of dirt obtained from Fig 3(b) using $c_q = 0.85$ are shown in Fig 5(b), which illustrates the substantial improvement in terms of detection accuracy to the extent that the dirt mask can now be directly associated with the dirt present in the original image in Fig 1(b). Besides, masks of dirt detected from Fig 1(a) and Fig. 1(c) are also given in Fig. 5.

IV. EXPERIMENTAL RESULTS

A. Visual Comparative Assessment

For comparison purposes we have considered established methodologies based on spatial, temporal and multistage median filtering (MMF). A detailed discussion of how these methods operate can be found in the Appendix. We denote the dirt masks detected using these methods as D_s, D_t, and D_m respectively.

(a) Spatial median filtering   (b) Temporal median filtering
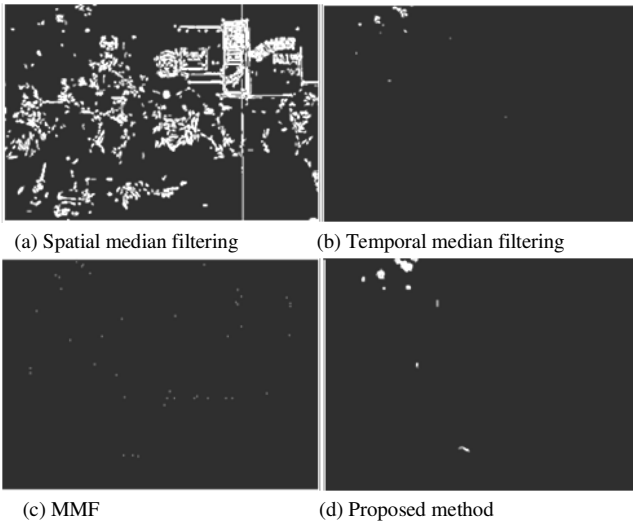
(c) MMF   (d) Proposed method

Fig. 6. Detected dirt masks using our method (d) at 90% confidence compared with results from (a) spatial, (b) temporal and (c) multistage median filtering.

Fig 6 shows comparisons with such spatial, temporal and multistage median filtering approaches. For spatial filtering, the parameters used were $r = 3$ and $t_s = 10$ while for temporal filtering $t_1 = 5$, $t_2 = 10$ and $t_3 = 15$. For MMF, five sub-windows are used for filtering and then thresholded by $t_m = 5$ (see Appendix). In temporal median filtering, we used the well-known Black-Anandan optical flow algorithm which yields dense motion fields of sub-pixel accuracy [6]. From Fig. 6 we can see that although small-area dirt can be detected by spatial and multistage median filtering to a reasonable extent, a lot of false alarms also occur. On the other hand, our method operating at a confidence level of 90% (see Fig 6d) yields comparable results to temporal median filtering while avoiding computationally expensive motion estimation overheads.

The next set of results involves the broadcast resolution (720x576) sequence "Pennine Way" which contains fast motion and textured background thus allowing for more demanding test conditions. Fig 7 shows detected dirt from one of the original frames with four enlarged partial images which contain the main areas of dirt in the scene, from which we can see that the result is reasonably accurate and relates well with a human viewer appraisal of the original image.
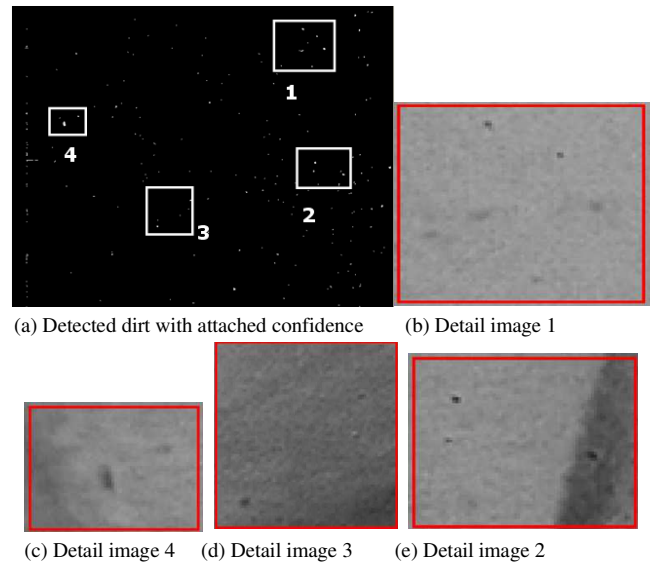

(a) Detected dirt with attached confidence   (b) Detail image 1

(c) Detail image 4   (d) Detail image 3   (e) Detail image 2

Fig. 7. Detected dirt with attached confidence (inverted for display) for Fig 1(d) from "Pennine Way" with four enlarged sub-images for details.

B. Quantitative Comparative Assessment

In the literature the objective assessment of dirt detection performance has occasionally involved the overlay of artificially-induced impairments on dirt-free images, such as in [1], [7] and [23]. While such an assessment framework is attractive from the point of view of ground-truth availability it is unfortunately unrealistic not least because the artificial generation of impairments is usually based on simplistic modeling that does not bear any relationship to actual dirt generation mechanisms.

Instead we have used manually-computed ground truth (GT)

maps from dirt-impaired images and we have employed two established criteria namely correct detection rate R_c and false alarm rate R_f. We assume that D_g is a ground truth dirt mask, that $D_x \in \{D_q, D_m, D_s, D_t\}$ i.e. that D_x is a detection mask obtained from any of the above methods (D_q refers to our method) then the above criteria are defined as follows:

$$R_c = \frac{Count(D_g \wedge D_x)}{Count(D_g)} \qquad (10)$$

$$R_f = \frac{Count(D_x \wedge \overline{D_g})}{Count(D_g)} \qquad (11)$$

where *Count* is a counter of the non-zero elements in a mask, and the operator $\wedge$ is the logical AND between two masks.

In this set of experiments we also used the "Pennine Way" sequence and chose the following parameter values for the methods under consideration: $r = 3$, $t_s = 15$, $t_1 = 25$, $t_2 = 15$, $\lambda = 4$, $s = 20$ and $c_q = 90\%$. Quantitative comparisons of detection performance using 20 consecutive frames are given in Fig 8. From these we can see that: (1) on average our method yields the highest successful detection rate and the lowest false alarm rate; (2) due to failure of motion estimation, temporal median filtering may occasionally yield very poor detection rates and frequent false alarms (i.e. between frames #205 and #211); (3) spatial median filtering yields a higher detection rate for some frames, however, this comes at the expense of more false alarms; (4) MMF achieves correct detection rate performance similar to that achieved by spatial median filtering but generates far fewer false alarms.

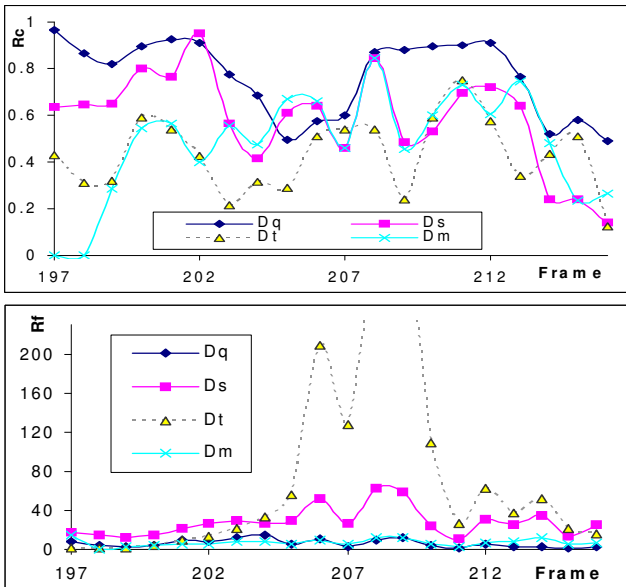Fig. 9 shows frame #210 of test sequence "Pennine Way"

Fig. 8. Quantitative comparison in terms of correct detection rate (top) and false alarm rate (bottom).

with dirt detected using our method as well as temporal median filtering. It is interesting to point out that false alarms occurring from the use of temporal median filtering are mainly due to inaccurate motion estimation (see Fig. 9c) while those occurring from the use of spatial median filtering and MMF (see Figs 6a and 6c) are due to sparkle type of noise or sharp edges. Owing to spatial segmentation and confidence-based weighting our method is less sensitive to both of those types of failure.

(a) Original frame image   (b) Dirt image of manual ground truth

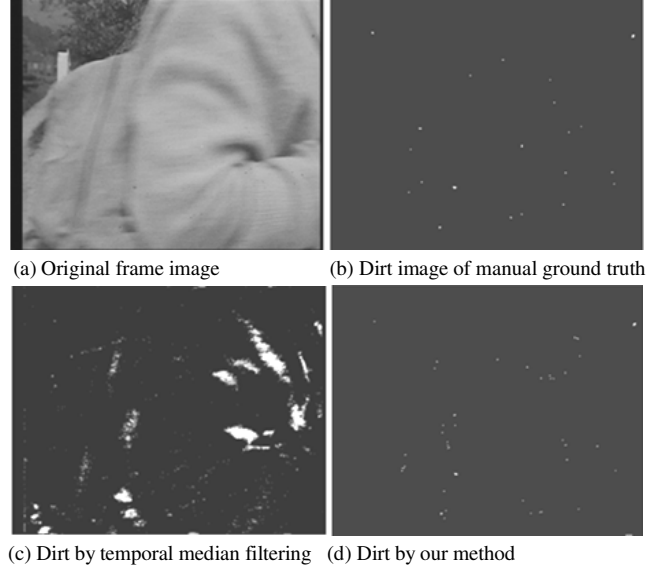(c) Dirt by temporal median filtering   (d) Dirt by our method

Fig. 9. One test image (a) and dirt image of ground truth (b) compared with detected dirt results using temporal median filtering (c), and our method (d).

In terms of computational complexity reflected by computer runtimes we noted that using a Pentium 4, 3.2GHz computing unit with 1GB of RAM, average processing times per frame of "Pennine Way" were 1.4, 53.8, and 6.2 seconds respectively for spatial, temporal and multistage median. For temporal median filtering, more than 99% of the computing time is spent for bi-directional motion estimation. Our method, on the other hand, only requires 0.9 seconds per frame, nearly 60 times faster than temporal median filtering.

V. PARAMETER INFLUENCE AND PERFORMANCE ASSESSMENT USING OBJECTIVE GROUND TRUTH DATA

In the proposed algorithm, there are three main parameters which impact performance, namely λ, s, and η. They respectively influence segmentation coarseness, dirt candidate selection and false-alarm suppression. In this section we assess how influential these parameters are and also compare performance with competing methods using objective ground truth. To the best of our knowledge, this is the first time that objective GT has been used in such a context. GT data were obtained by using infrared scanning. Due to their non-transparency, dirt areas always appear dark in the generated GT data. Partially transparent scratches and other artefacts may also appear in lighter grey shades. Taking this into account we inverted the intensity of the original GT data and then

thresholded it to obtain a binary mask. Fig 10 shows part of the luminance component of one original frame and the corresponding greylevel GT data and binary mask of dirt.

(a) Y component　　　(b) Objective GT　　　(c) dirt mask of GT

Fig. 10. (a) luminance component of original frame, (b) corresponding greylevel GT data and (c) binary GT mask obtained from (b).

A. Influence of λ

In Section II (A) we mentioned that segmentations obtained from larger λ seem to be more accurate in identifying dirt pixels. We illustrate this in Fig. 11 by computing correct detection and false alarm rates using four different values of λ with the other parameters set at $s = 150$ and $\eta = 2.5$.

Due to the fact that a higher λ will inevitably cause over-segmentation this will further lead to a higher correct detection rate but also more false alarms and vice versa. By setting λ to values higher than 5 we have observed no obvious improvement in detection rates at the expense of more false alarms.
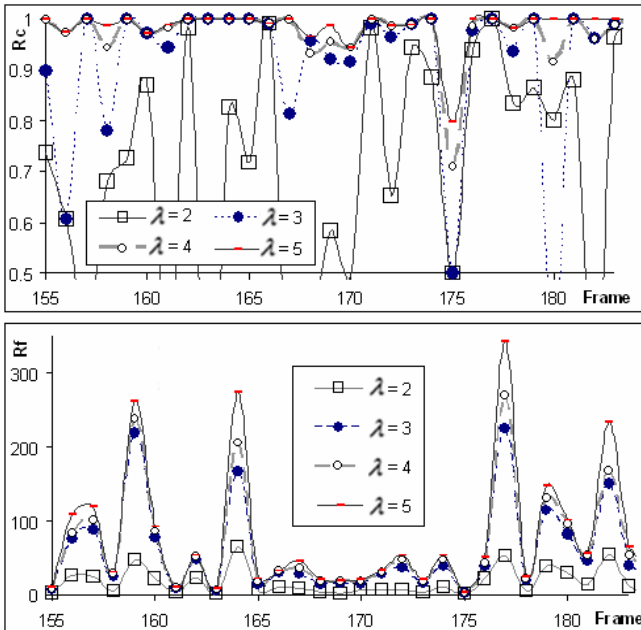

Fig. 11. Correct detection rate (top) and false alarm rate (bottom) using different λ with $s = 150$ and $\eta = 2.5$.

B. Influence of s

In Fig. 12 we illustrate the influence of thresholding according to region size using $\lambda = 6$ and $\eta = 2.5$. We can see that a higher value of s, for example $s > 150$, is necessary to ensure a sufficiently high detection rate. However, our experimental evidence also suggests that higher threshold values tend to generate more false alarms.
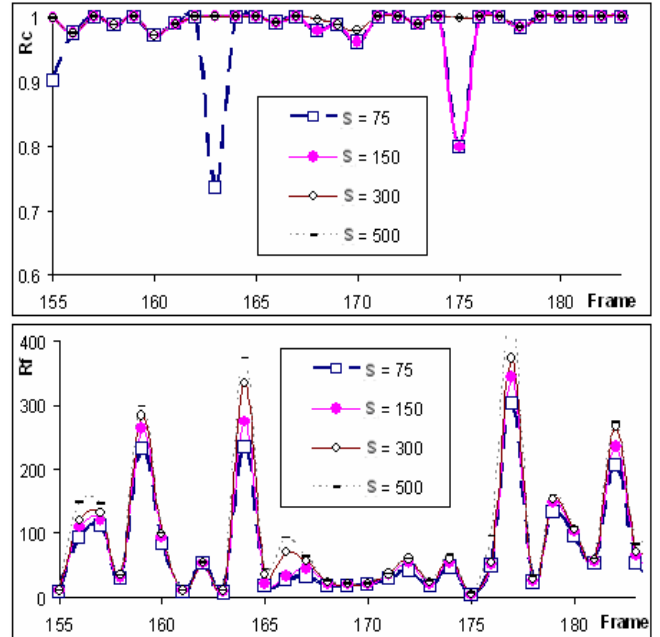

Fig. 12. Correct detection rate (top) and false alarm rate (bottom) using different s with $\lambda = 6$ and $\eta = 2.5$.
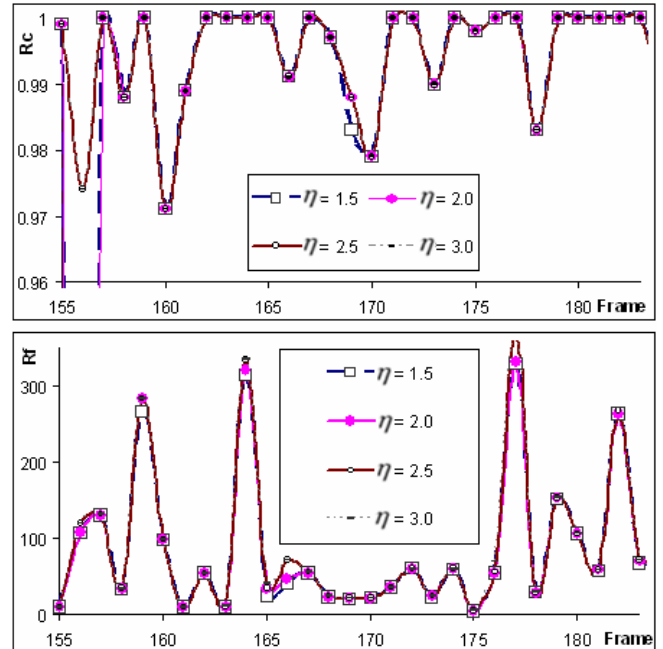

Fig. 13. Correct detection rate (top) and false alarm rate (bottom) using different η with $\lambda = 6$ and $s = 300$.

C. Influence of η

In Fig. 13 we illustrate the influence of parameter η which controls false alarm suppression. We can see that for over-segmented images (obtained by using high λ value) there is no appreciable benefit to be gained by η variations with respect to both correct detection rate and false alarm rate. This underlines that our algorithm is relatively insensitive to the choice of parameter η under those conditions.

D. Comparative Performance Assessment

Finally in Fig 14 we compare the performance of our algorithm against spatial median filtering and the ROD (rank ordered difference) detector [5] with respective parameter values $r = 3$, $t_s = 5$, $t_1 = 5$, $t_2 = 10$, and $t_3 = 15$. We considered ROD because it offers the best correct detection rate performance relative to the other two motion-compensated methods, i.e. temporal median filtering and SDIp. We have also avoided including MMF in our comparison framework as it proved fairly inefficient for the type of test material under consideration.
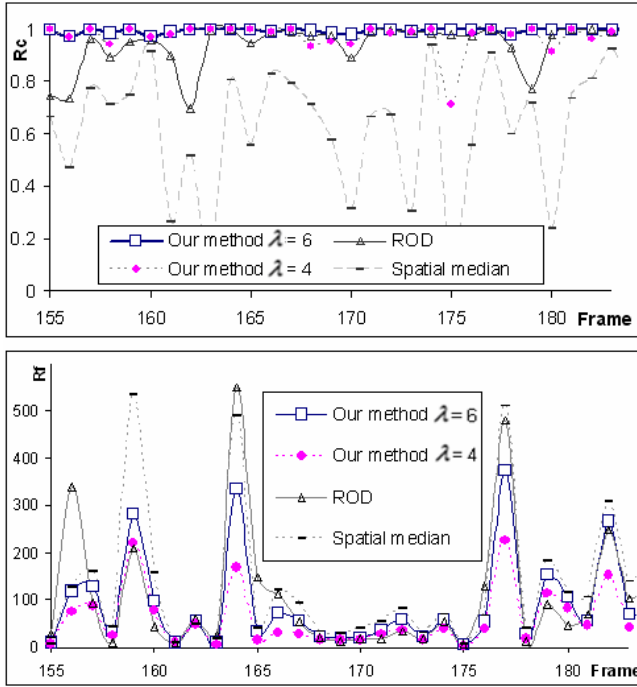

Fig. 14. Correct detection rate (top) and false alarm rate (bottom) using different detection methods, in two of our results $\eta = 2.5$ and $s = 250$.

From Fig. 14 we can see that the proposed algorithm offers better performance in terms of correct detection rate. Moreover, on average, our method generates a similar level of false alarms relative to ROD while avoiding the additional complexity associated with the use of motion estimation and compensation. On the other hand we can see that spatial median filtering offers inferior performance both in terms of correct detection rate as well as false alarm rate.

VI. CONCLUSIONS

We have demonstrated that image segmentation can provide a useful framework for dirt detection in film restoration. One advantage of our approach in relation to techniques operating in the spatial domain is the lack of any requirement related to the specification of a filter such as shape and size. Another relative advantage is the capability of our approach to detect dirt as connected clusters of pixels in the shape of a region which is consistent with actual manifestations of dirt. A key advantage relative to motion estimation methods is the much lower computational complexity. Finally, confidence weighting derived from interframe information allows fine tuning according to operator preference. Experimental evidence obtained using both manually-derived and objective ground truth data suggests that the proposed method is effective, robust and outperforms conventional spatial, temporal and multistage median filtering.

APPENDIX

In dirt detection based on spatial median filtering, pixels are processed as follows. For each pixel (i, j) in the current frame f_n, a window W of radius r is defined as

$$W(i, j, r) = \{(i_1, j_1); |i_1 - i| \le r, |j_1 - j| \le r\} \quad (12)$$

The total number of pixels in W is $(2r + 1)^2$. Pixel values in W are sorted in monotonically increasing or decreasing order and the median value is taken as the filter output. If g_n is the filter output i.e. $g_n(i, j) = median(f_n(i', j'))$, where $(i', j') \in W(i, j, r)$, then the dirt detected by spatial median filtering (relative to a threshold t_s) is given by

$$D_s(i, j) = \begin{cases} 1 & if \quad |g_n(i, j) - f_n(i, j)| > t_s \\ 0 & otherwise \end{cases} \quad (13)$$

For temporal median filtering, at least three frames are needed: the current frame f_n and the two motion-compensated neighboring frames C_{n-} and C_{n+}. In [2] the detected dirt D_t is defined by Eq. (14) with thresholds, t_1 and t_2, satisfying $t_1 > t_2$.

$$D_t(i, j) = \begin{cases} & |C_{n-}(i, j) - f_n(i, j)| > t_1, \\ 1 & if \quad |C_{n+}(i, j) - f_n(i, j)| > t_1, \\ & |C_{n-}(i, j) - C_{n+}(i, j)| < t_2 \\ 0 & otherwise \end{cases} \quad (14)$$

Regarding multistage median filtering (MMF), the bi-directional version was considered for our experiments [10]. The basic concept of MMF is based on the five sub-windows defined in three consecutive frames (see Fig 15). After filtering

using Eq. (15)-(17), the dirt mask, D_m, is determined in the same way as in the spatial median filtering approach above.

$$z_l = med[w_l] \qquad l \in \{0,1,2,3,4\} \tag{15}$$

$$z_{max} = max[z_l], z_{min} = min[z_l] \quad l \in \{1,2,3,4\} \tag{16}$$
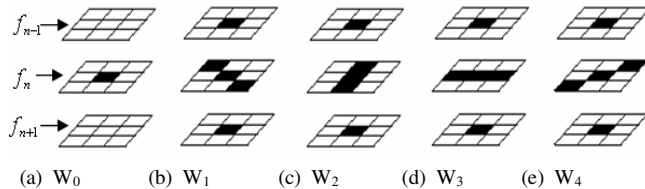
$$MMF(z) = med[z_{max}, z_{min}, z_0] \tag{17}$$

(a) W_0　　(b) W_1　　(c) W_2　　(d) W_3　　(e) W_4

Fig. 15. Sub-windows defined in three consecutive frames for bi-directional MMF (radius = 1).

REFERENCES

[1] A.C. Kokaram, "On missing data treatment for degraded video and film Archives: a survey and a new Bayesian approach," *IEEE Trans. Image Proc.,* vol. 13-3, pp. 397-415, 2004

[2] P. Schallauer, A. Pinz, and W. Haas, "Automatic restoration algorithms for 35mm film," *J. Computer Vision Res.,* vol. 1-3, pp. 59-85, 1999.

[3] A. Gangal, T. Kayikcioglu, and B. Dizdarogle, "An improved motion-compensated restoration method for damaged color motion picture films," *Signal Processing: Image Comm.*, vol. 19, pp. 353-368, 2004.

[4] A.C. Kokaram, *Motion Picture Restoration*, Berlin, Germany: Springer-Verlag, 1998.

[5] M.J. Nadenau and S.K. Mitra, "Blotch and scratch detection in image sequences based on rank ordered differences," in *Proc. of 5th Int. Workshop on Time-Varying Image Processing*, Florence, 1996, pp. 27-35.

[6] M.J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *J. CVIU,* vol. 63-1, pp. 75-104, 1996.

[7] 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 Trans. Cir. Sys. for Video Tech.,* vol. 13-5, pp. 406-416, 2003.

[8] D. Krishnan, M. N. Chong, and S. Kalra, "On the computational aspects of Gibbs-Markov random field modelling of missing data in image sequences," IEEE Trans. Image Proc., vol. 8-8, pp. 1139-1142,1999.

[9] A. Nieminen, P. Heinonen, and Y. Neuvo, "A new class of detail-preserving filters for image processing," *IEEE Trans. PAMI*, vol. 9, pp. 74-90, 1987.

[10] G. Arce, "Multistage order statistic filters for image sequence processing," *IEEE Trans. Signal Proc.,* vol. 39, pp. 1146-1163, 1991.

[11] H. Senel, R. Peters, and B. Dawant, "Topological median filters," *IEEE Trans. Image Proc.*, vol. 11, pp. 89-104, 2002.

[12] R. Hardie and C. Boncelet, "LUM filters: a class of rank-order-based filters for smoothing and sharpening," *IEEE Trans. Signal Proc.,* vol. 41, pp. 1061-1076, 1993.

[13] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. PAMI*, vol. 16-6, pp. 641-647, 1994.

[14] M. K. Özkan, M. I. Sezan, and A. M. Tekalp, "Adaptive motion-compensated filtering of noisy image sequences," IEEE Trans. Cir. Sys. For Video Techno., vol. 3-4, pp. 277-290, 1993.

[15] R. Storey, "Electronic detection and concealment of film dirt," *SMPTE J.*, pp. 642-647, June 1985.

[16] R. D. Morris, "Image sequence restoration using Gibbs distribution," PhD. Dissertation, Cambridge University, Cambridge, U. K., 1995

[17] R. D. Morris and W. J. Fitzgerald, "Replacement noise in image sequences, detection and interpolation by motion field segmentation," in *Proc. ICASSP*, vol. V, pp. v/245-v/248, 1994.

[18] A. T. Erdem, M. I. Sezan, and M. K. Ozkan, "Motion-compensated multiframe Wiener restoration of blurred and noisy image sequences," in *Proc. IEEE ICASSP*, vol. 3, pp. 293-296, 1992.

[19] S.Geman and D.Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. PAMI*, vol. 6, pp. 721-741, 1984.

[20] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. PAMI*, vol. 14, pp. 367-383, 1992.

[21] A. C. Kokaram, R. Bornard, A. Rares, D. Sidorov, J.-H. Chenot, L. Laborelli, and J.Biemond, "Robust and automatic digital restoration systems: coping with reality," in Proc. Int. Broadcasting Convention, pp. 405-411, 2002

[22] A. Rares, M. J. T. Reinders, and J. Biemond, "Complex event classification in degraded image sequences," in *Proc. ICIP*, vol. 1, pp. 253-256, 2001.

[23] A. C. Kokaram, R. Morris, W. J. Fitzgerald, and P. J. W. Rayner, "Detection of missing data in image sequences," *IEEE Trans. Image Proc.*, vol. 40-11, pp. 1496-1508, 1995.

[24] M. N. Chong and D. Krishnan, "An edge-preserving MRF model for the detection of missing data in image sequences," *IEEE Signal Proc. Letters*, vol. 5-4, pp. 81-83, 1998.

[25] S. Kalra, M. N. Chong, and D. Krishnan, "A new auto-regressive (AR) model-based algorithm for motion picture restoration," in *Proc. ICASSP*, vol. 4, pp. 2557-2560, 1997.

[26] A. C. Kokaram and S. J. Godsill, "MCMC for joint noise reduction and missing data treatment in degraded video," IEEE Trans. Signal Proc., vol. 50-2, pp. 189-205, 2002.

[27] A. C. Kokaram and P. J. W. Rayer, "Removal of replacement noise in motion picture sequences using 3D autoregressive modelling," *Signal Processing*, vol. 7, pp. 1760-1763, 1994.

[28] O. Buisson, B. Besserer, S. Boukir, and F. Helt, "Deterioration Detection for Digital Film Restoration," in Proc. CVPR, pp. 78-84, 1997.

[29] E. D. Ferrandière, *Mathematical Morphology and Motion Picture Restoration*, New York: Wiley, 2001.

[30] T. S. Huang and Y. P. Hsu, "Image sequence enhancement," in *Image Sequence Analysis*, T. S. Huang, Ed. Berlin, Germany: Springer-Verlag, pp. 289-309, 1981.

[31] E. Dubois and S. Sabri, "Noise reduction in image sequences using motion-compensated temporal filtering," *IEEE Trans. Commu*, vol. 32-7, pp. 826-831, July 1984.

[32] T. J. Dennis, "Nonlinear temporal filter for television picture noise reduction," *IEE Proc.*, part G, vol. 127, pp. 52-56, 1980.

[33] R. H. Mcmann, S. Kreinik, J. K. Moore, A. Kaiser, and J. Rossi, "A digital noise reducer for encoded NTSC signals," *SMPTE J.*, vol. 87, pp. 129-133, 1979.