DIRT DETECTION FOR ARCHIVE FILM RESTORATION USING AN ADAPTIVE SPATIO-TEMPORAL APPROACH

Jinchang Ren, Theodore Vlachos

Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, GU2 7XH, United Kingdom
{j.ren,t.vlachos}@surrey.ac.uk

Keywords: Dirt detection, spatio-temporal filtering, film restoration

Abstract

We propose an adaptive spatio-temporal approach for the detection of dirt in archive film sequences. By combining useful features from conventional stand-alone spatial and temporal approaches our method achieves better performance in terms of both correct dirt detection rate and false alarm rate. Furthermore the incorporation of confidence weighting offers a useful control mechanism for semi-automatic restoration scenarios involving manual operator intervention. Overall our scheme offers a practical solution to the dirt detection problem without additional computational complexity overheads.

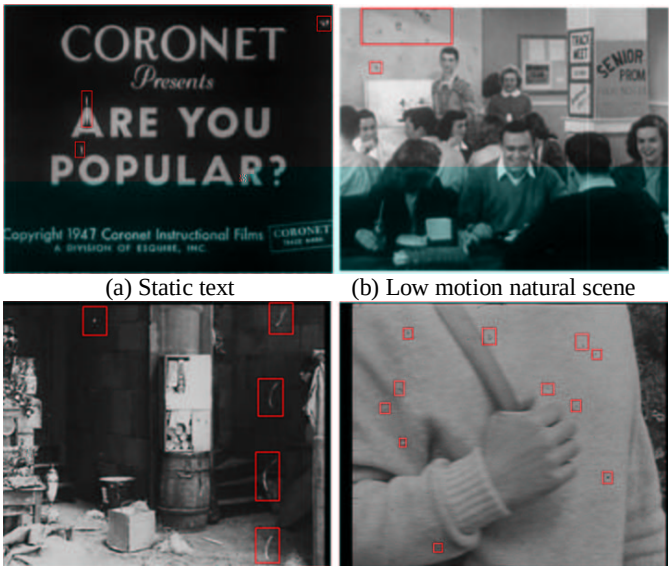
1 Introduction

In recent years, the emergence of new multimedia and broadcasting outlets is dramatically improving public access to cultural assets of unique educational and entertainment value. By improving baseline picture quality and by reducing the perceptual impact of archive related impairments, film archive restoration (FAR) can meet viewers' aesthetic expectations and enrich the viewing experience thereby facilitating access to and appreciation of historical film footage.

Consequently, FAR has attracted lots of interest from many broadcasters and media holders while several high-profile collaborative projects have received EU (European Union) funding such as AURORA (Automatic Restoration of Original Film and Video Archives, 1995), BRAVA (Broadcast Restoration of Archives by Video Analysis, 1999) and more recently PrestoSpace (Preservation towards Storage and access Standardised Practices for Audiovisual Contents in Europe, 2004).

In their lifetime, films may suffer impairments due to environmental hazards such as humidity and dust, chemical instabilities, improper storage and handling practices and even poorly maintained projectors [6, 9, 12]. The suppression of these impairments has vital implications on the efficiency of video coding algorithms, such as MPEG (1,2,4) and AVC, used in many broadcast and multimedia distribution chains, including digital television, video on demand and Internet-based online cinema.

In this paper, we focus on the detection of impairments occasionally referred to 'dirt'. These are among the most commonly encountered artefacts and hence their successful detection is a priority issue in any archive restoration system [9]. In general, dirt is a temporally impulsive (single-frame) event, appearing mostly as dark or bright opaque spots of random size, shape and location. It is due to particles which are attached to the film or localised abrasions which occurred during storage or when the film passed through various transport mechanisms [6, 9]. Fig 1 gives some examples of dirt which are shown bounded by red rectangles.



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

(c) Moderate motion natural scene (d) Fast motion natural scene

Figure 1: Examples of dirt

Dirt detection typically involves two generic steps, namely the identification of inconsistency of a pixel in relation to its spatio-temporal neighbourhood followed by thresholding. The first step is conventionally implemented by a suitable combination of intra/inter type of filtering. In the second step a well-established principle is that the choice of threshold will influence the balance between false alarm rate and correct detection rate.

According to how inconsistency is determined, we can categorize current dirt detection methods into three classes,

i.e. spatial filtering, temporal filtering and spatial-temporal filtering.

In spatial filtering, dirt is viewed as a spatial impulse and is typically detected using neighbourhood information in the current frame. Prominent among spatial detection methods are those featuring median filtering [7, 11, 13]. Nieminen *et al* [11] presented a multi-level median filter (MLF) to reduce the influence of outlier values while preserving edges. Their filter firstly calculates separate median values for horizontal, vertical, and two diagonal transects in a given window, and then the minimum and maximum of these four values are found. The median of the minimum, maximum, and original raster value in the central of the window is taken as the output of the filter. Hardie, and Boncelet [7], proposed LUM (lower-upper-middle) filters, in which two parameters are utilised for adjustable smoothing and sharpening of images. Senel *et al* [13] proposed a topological median filter to extract edges in noise; however, the filtered images are of unacceptable visual quality in most cases.

In temporal filtering approaches, dirt is viewed as a temporal impulse (single-frame incident) and hence treated by inter-frame processing by taking into account at least three consecutive frames [8, 12]. Storey's work was perhaps the earliest contribution to the electronic detection and concealment of dirt [14]. In his hardware-based system, a pixel was flagged as dirt if the corresponding absolute differences between the current frame and each of the previous and next frames were high. Kokaram extended this idea using motion-compensated differences [8], in which the so-called "Spike Detection Index" (SDI) was proposed. The basic SDI detector was based on the identification of high absolute differences between the current frame and two compensated images, and an expanded SDI detector, SDIp, additionally required sign consensus of the two differences above. In Schallauer *et al* [12], a pixel is taken as dirt and filtered if both its absolute differences between current frame and the two compensated images exceed a first (higher) threshold while at the same time the absolute difference between the two compensated images is less than a second (lower) threshold.

In spatio-temporal filtering, pixel inconsistency is determined by the examination of both spatial and temporal neighbourhoods. In [2], Arce applied a multi-stage order statistic filter approach (MOS) extending MLF to three consecutive frames, to image sequence noise suppression. At the same time, Arce also proposed a three-frame LUM variant for image smoothing purposes. In [1], Alp *et al* introduced the so-called ML3D algorithm, in which two groups of windows are defined in three frames and their median values are determined. Then, the median of the two median values, herein which is calculated as the average of the two median values, is taken as the output. Nadenau, and Mitra [10], have used a rank order detector (ROD), in which a total of seven pixels from three consecutive frames (with motion compensation) are compared against three thresholds. Gangal *et al* extended ROD to five frames to improve accuracy in heavily corrupted images or occluded blotches [5]. In [8], Kokaram presented an extended version of ML3D, ML3Dex, which applied ML3D filtering to the output of the ROD detector. In [6], Hamid *et al* employed non-motion-compensated soft morphological filtering (SMF) in three consecutive frames and argued that LUM and ML3Dex are not effective towards restore fast-moving objects in image sequences. However, SMF seems impractical for most applications because it needs sufficient representative samples of dirt for training to optimise the size and shape of the filters.

Generally speaking, although motion compensation can potentially become an essential component of a dirt detection algorithm, it is well known that it does not degrade gracefully when motion estimation fails and may thus generate unpredictable results. For such a reason spatial filtering may be regarded as a useful tool either as a complement to motion-compensated approaches or even as an alternative to them in the framework of spatio-temporal treatment that takes into account exclusively raw (non-motion-compensated) frame differences. Current algorithms do not combine the above concepts very efficiently. 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 perform poorly when motion cannot be accurately estimated.

In this paper we develop an adaptive approach for dirt detection using a combination of spatial and temporal filtering. Furthermore, confidence weighting is incorporated offering additional flexibility in automatic and operator-assisted restoration scenarios as it allows a variable degree of treatment according to preference.

This paper is organized as follows. In Sections 2 and 3, conventional stand-alone methods for dirt detection with and without motion compensation respectively are discussed. Section 4 contains a description of the proposed adaptive method. Experimental results are given in Section 5 and conclusions are drawn in Section 6.

2 Non-motion compensated dirt detection

Spatial median filtering is a basic tool for dirt detection when motion compensation is unavailable. In this section, we will discuss three conventional methods, namely standard spatial median filtering (SSMF) [4], LUM [6] and MOS [2]. The first two approaches are exclusively spatial while the third is spatial-temporal.

2.1 Definitions

In SSMF and LUM for each pixel (i, j) in the current frame f_n, a window W of radius r is defined as

$$W(i, j, r) = \{ f(i_1, j_1) \}, | i_1 - i | \le r, | j_1 - j | \le r \qquad (1)$$

Hence the total number of pixels in W is $N = (2r+1)^2$. For odd N, we define $N_0 = (N+1)/2$.

We denote $W = \{x_1, x_2, ..., x_N\}$, and the rank-ordered set is given by

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(N)} \qquad (2)$$

The central pixel in the original current frame and filtered image are denoted as x' and y', respectively. In SSMF, we simply have

$$y' = x_{(N_0)} \tag{3}$$

In LUM, two parameters, k and l, are introduced for smoothing and sharpening, respectively. Typically it holds that $1 \le k \le l \le N_0$. Then, the filtered output is defined as

$$y' = \begin{cases} x^L, & if \quad x' \le (x^L + x^U)/2 \\ x^U, & otherwise \end{cases} \tag{4}$$

where x^L and x^U are the corresponding outputs of the smoothing and sharpening processes, which are given by

$$x^L = median\{x_{(k)}, x', x_{(l)}\} \tag{5}$$

$$x^U = median\{x_{(N-k+1)}, x', x_{(N-l+1)}\} \tag{6}$$

An alternative definition of LUM filtering is given in (7), in which $t_l = (x_{(l)} + x_{(N-l+1)})/2$.

$$y' = \begin{cases} x_{(k)}, & if \quad x' < x_{(k)} \\ x_{(l)}, & if \quad x_{(l)} < x' \le t_l \\ x_{(N-l+1)}, & if \quad t_l < x' < x_{(N-l+1)} \\ x_{(N-k+1)}, & if \quad x_{(N-k+1)} < x' \\ x', & otherwise \end{cases} \tag{7}$$

In Arce [2], LUM was further applied to a $3\times3\times3$ spatial-temporal window. We denote his LUM version as $LUM'(N,k)$, where we have $N = 27$ and $k \le 14$. The output of the filter is given by

$$y' = median(x_{(k)}, x', x_{(N-k+1)}) \tag{8}$$

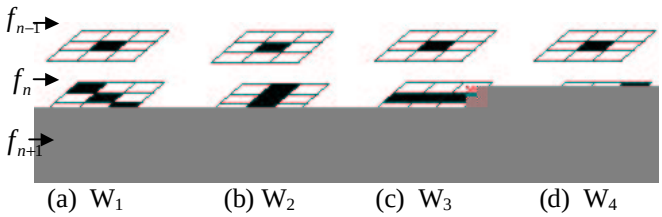
Regarding MOS, the bi-directional variant used in our experiments [2] is defined on the basis of four sub-windows in three consecutive frames as shown in Fig 2. Firstly the median of each sub-window is obtained, and then the four median values are taken for further filtering.

$$z_l = median[w_l] \qquad \forall l \in [1,4] \tag{9}$$

$$z_{\max} = \max[z_l] \tag{10}$$

$$z_{\min} = \min[z_l] \tag{11}$$

$$y' = median[z_{\max}, z_{\min}, x'] \tag{12}$$



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

Figure 2: Sub-windows defined in three frames for bi-directional MOS filtering (radius = 1).

2.2 Dirt detection and analysis

Let g_n be the output image after filtering, i.e. $g_n(i,j) = y'(i,j)$. Dirt D is then detected as

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

where t_s is a predefined threshold.

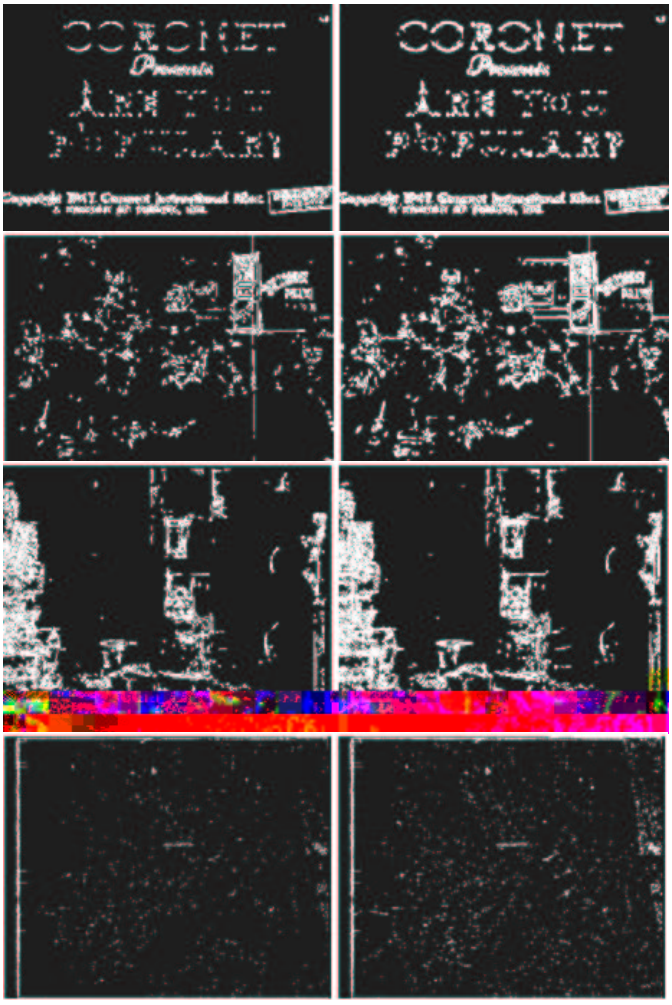


Figure 3: Dirt detected from Fig 1 by SSMF with $t_s = 10$ and window of 5×5 (left) and 7×7 (right), respectively.

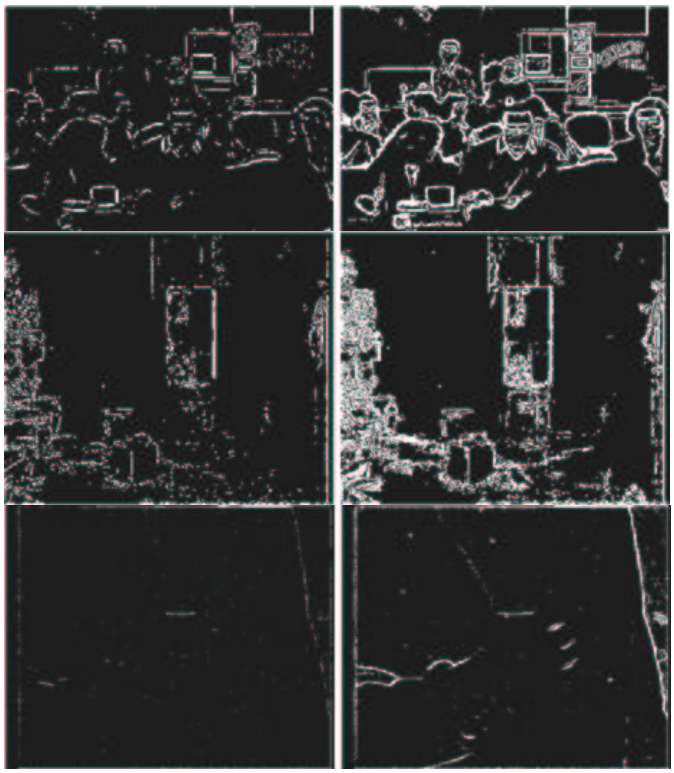


Figure 4: Dirt detected from Fig 1 by $LUM(5,10)$ with $t_s = 10$ and window of 5×5 (left) and 7×7 (right), respectively.

With SSMF and LUM, detection performance is very sensitive to the size and shape of window W. A square window in usually adopted when we have no a priory information suggesting otherwise. With reference to the original image in Fig 1, Fig 3 and Fig 4 show detected dirt using SSMF and LUM using different window sizes, where $t_s = 10$, $k = 5$, $l = 10$ and $r = 2,3$, respectively.

Comparing Fig 1, Fig 3 and Fig 4, we can conclude that: 1) SSMF performs better than LUM; 2) the false alarm rate especially near sharp edges or sparkle/noise is unacceptably high; 3) LUM has less of a smoothing effect than SSMF and consequently preserves better original features; 4) larger windows are capable of detecting larger areas of dirt.

In Fig 5, we show the dirt detection results using MOS and $LUM'(27,9)$, respectively. Owing to temporal information, most of the false alarms that occurred from processing Fig 1(a) and Fig 1(b) are avoided. However, a number of false alarms remain for Fig 1(c), while additional ones occurred for Fig 1(d).

Moreover, it seems that LUM' is superior to MOS in relation to correct dirt detection rate, but also generates more false alarms. The application of SSMF to fast moving sequences, such as Fig 1(d), causes less false alarms than LUM'. In section 5, these two methods will be further evaluated.

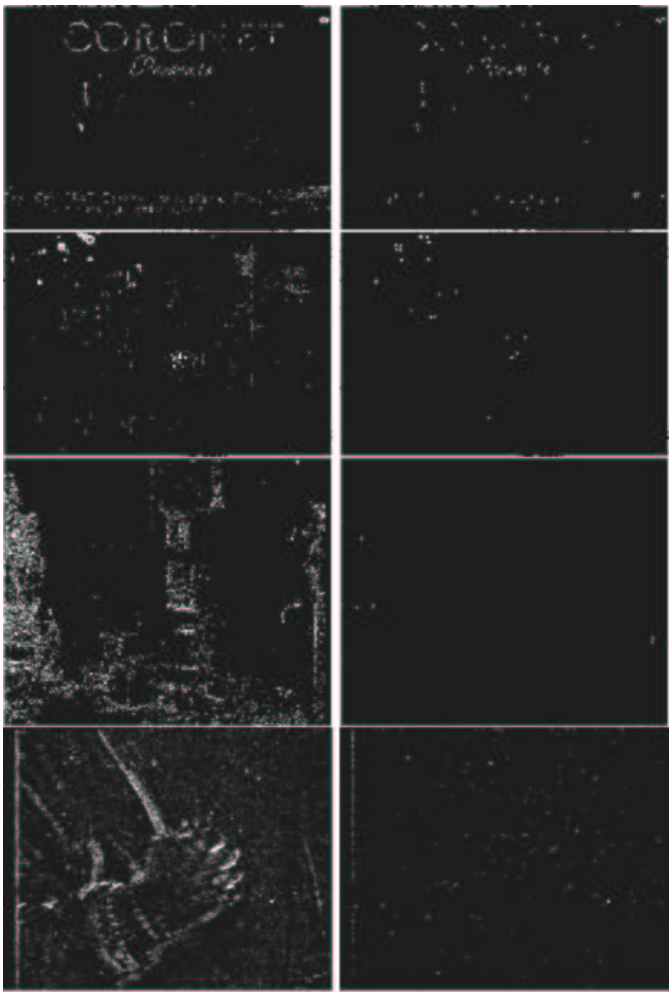


Figure 5: Dirt detected from Fig 1 by $LUM'(27,9)$ (left) and MOS (right) with $t_s = 10$.

3 Motion compensated dirt detection

Motion compensation is typically combined with ttemporal median filtering. In this case, at least three frames are needed: the current frame f_n and the two motion compensated frame neighbours, C_{n-} and C_{n+}. We define D_{n-} and D_{n+} as the differences between each of these two images and the current frame:

$$D_{n-} = C_{n-} - f_n \quad (14)$$

$$D_{n+} = C_{n+} - f_n \quad (15)$$

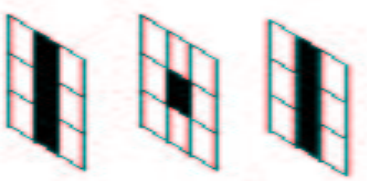
Using the above notation we subsequently consider established approaches such as double-threshold temporal median filtering (DTMF) [12], SDIp [8] and ROD [10].

In [12], a pixel is declared as dirt if it satisfies (16) where t_1 and t_2 are two given thresholds with $t_2 > t_1$.

$$D_t(i,j)=\begin{cases}1 & if \quad |D_{n-}(i,j)|>t_2, |D_{n+}(i,j)|>t_2, \\ & \quad |C_{n-}(i,j)-C_{n+}(i,j)|<t_1 \\ 0 & otherwise\end{cases} \quad (16)$$

In [8], dirt pixels are defined as

$$D_{SDIp}=\begin{cases}1 & if \quad |D_{n-}(i,j)|>t_1, |D_{n+}(i,j)|>t_1, \\ & \quad D_{n-}D_{n+}>0 \\ 0 & otherwise\end{cases} \quad (17)$$



(a) C_{n-} (b) f_n (c) C_{n+}

Figure 6: Window structure for spatial-temporal filtering in ROD for dirt detection.

To determine dirt in ROD, three pair of pixels are extracted from C_{n-} and C_{n+} (see Fig 6). These six pixels are sorted in increasing order in a list $[r_1, r_2, ..., r_6]$ where r_6 is maximum. Then the median of the list is extracted as $med=(r_3+r_4)/2$. If $f_n(i,j)>med$, we define $e_k = f_n(i,j)-r_{7-k}, k\in[1,3]$, otherwise $e_k = r_k - f_n(i,j)$. Dirt is then detected if we have any e_k greater than t_k (18).

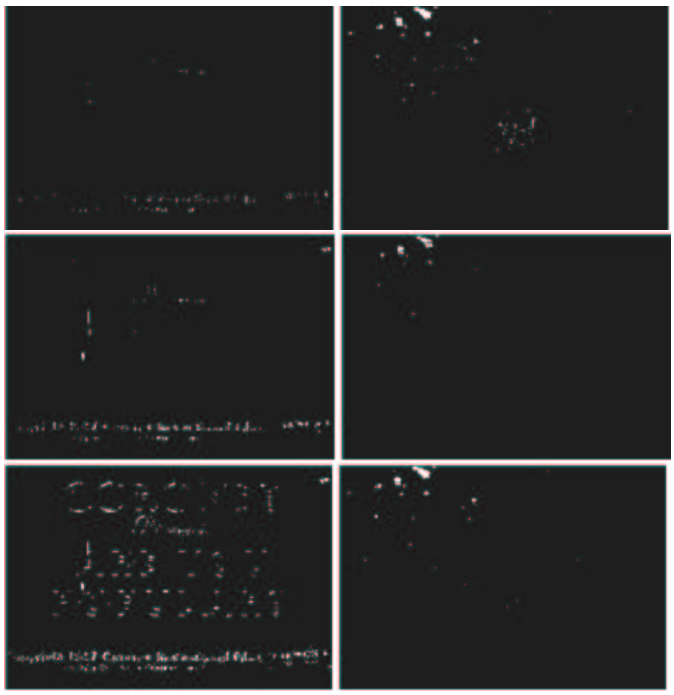


Figure 7: Dirt detected from low motion material (Fig 1a and 1b) by DTMF (top), SDIp (middle) and ROD (bottom) with thresholds $t_1=15, t_2=25$ and $t_3=35$.

$$D_{ROD}=\begin{cases}1 & if \quad e_1>t_1 \quad or \quad e_2>t_2 \\ & \quad or \quad e_3>t_3 \\ 0 & otherwise\end{cases} \quad (18)$$

Fig 7 and Fig 8 shows detected dirt using the three methods above with $t_1=15, t_2=25$ and $t_3=35$. Motion compensation was implemented by using dense motion fields of sub-pixel accuracy from the well-known Black-Anandan optical flow algorithm [3]. From the two figures we can see that dirt detected from ROD has a worse false alarm rate than that from the other two methods. On the other hand it is more accurate regarding the identification of the actual shape of a dirt cluster. It should also be noted that most of these false alarms are caused by failure of motion compensation.

DTMF and SDIp generate comparable results regarding correct detection and false alarms, but their performances are sensitive to the availability of dirt-free background. SDIp performs better for Fig 1(a) but worse for Fig 1(b) and Fig 1(d). DTMF and SDIp perform similarly for Fig 1(c).

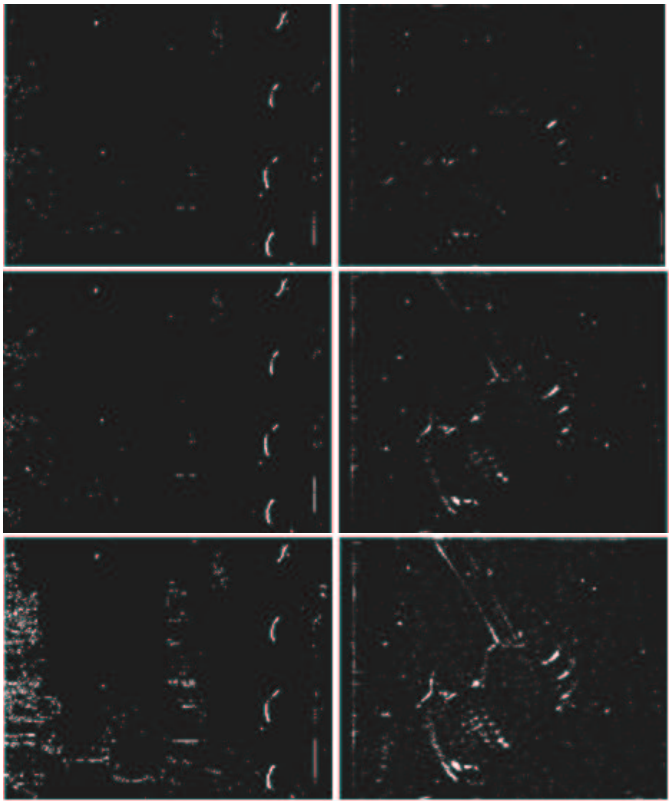


Figure 8: Dirt detected from medium and fast motion sequences (Fig 1c and 1d) by DTMF (top), SDIp (middle) and ROD (bottom) with thresholds $t_1=15, t_2=25$ and $t_3=35$.

4 Adaptive spatial-temporal detection with attached confidence

As already stated while motion-compensation contributes significantly accurate dirt detection, it also causes many false alarms when motion estimation fails. In such cases we

propose an adaptive fallback strategy based on spatial filtering which is explained below.

4.1 Adaptive spatio-temporal filtering

For a given frame f_n, let C_{n-}, C_{n+} and g_n be the two motion compensated images and spatially filtered image, respectively. We define E_{n-}, E_{n+} and E_s as motion-compensated and spatial filtering residuals as follows:

$$E_{n-}(i,j) = |C_{n-}(i,j) - f_n(i,j)| \quad (19)$$

$$E_{n+}(i,j) = |C_{n+}(i,j) - f_n(i,j)| \quad (20)$$

$$E_s(i,j) = |g_n(i,j) - f_n(i,j)| \quad (21)$$

Let μ_{n-}, μ_{n+} and μ_s be the mean-square values of E_{n-}, E_{n+} and E_s, respectively. These are computed over localised windows. Low valued μ_{n-} and/or μ_{n+} imply reliable forward and/or backward motion compensation. On the other hand, high values of these parameters suggest motion estimation failure and call for a spatial filtering approach.

Let $\mu_n = h(\mu_{n-}, \mu_{n+})$ provide a combination of μ_{n-} and μ_{n+} suitable for a threshold type of decision so that, for example, motion-compensation is preferred if $\mu_n \le \mu_s$, and vice-versa. Function h is given by

$$h(x,y) = \lambda(x,y)\frac{xy}{x+y} \quad (22)$$

where $\lambda \in (0,2]$ is a parameter controlling motion estimation influence. The proposed adaptive spatial-temporal approach is summarised below:

1) For a given frame f_n, calculate E_{n-}, E_{n+} and E_s from C_{n-}, C_{n+} and g_n, respectively;
2) Partition f_n, E_{n-}, E_{n+} and E_s into $N \times N$ blocks and for each block in E_{n-}, E_{n+} and E_s compute μ_{n-}, μ_{n+} and μ_s as well as μ_n;
3) If $\mu_n \le \mu_s$, motion-compensated methods is applied to the corresponding block in f_n for dirt detection, otherwise, spatial filtering method is employed;
4) Repeat until all blocks have been processed.

4.2 Confidence weighting

The assumption that dirt is a single-frame event leads naturally to the idea of using inter-frame information. 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 raw forward and backward frame differences, respectively. Then, we define d_n as:

$$d_n = \begin{cases} \dfrac{2d_{n-}d_{n+}}{|d_{n-}| + |d_{n+}|} & if \quad d_{n-}d_{n+} > 0 \\ 0 & otherwise \end{cases} \quad (23)$$

This attains its maximum value when an idealised 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).

For each value m in d_n, dirt probability is defined as:

$$p_n(m) = \lambda_0^{-1} \int_{m_0+1}^{m} p_d(x)dx \quad (24)$$

$$\lambda_0 + \int_0^{m_0} p_d(x)dx = 1 \quad (25)$$

where p_d is the intensity PDF (probability density function) of d_n, which can be derived from the histogram of d_n. Parameter λ_0 is used to normalise $p_n(m)$ within [0,1], and m_0 to control the removal of static background.

It is worth noting that $m_0 = 0$ in Eq. (24) amounts to histogram equalisation of d_n. Nevertheless, a static background in the three consecutive frames, may force most pixel values in d_n near zero therefore straightforward histogram equalisation is not useful in this context.

Let μ, γ and σ be the mean, median and variance of the distribution of values in d_n, and let m_0 be determined by

$$m_0 = \frac{\mu + \gamma}{2} + \sigma \quad (26)$$

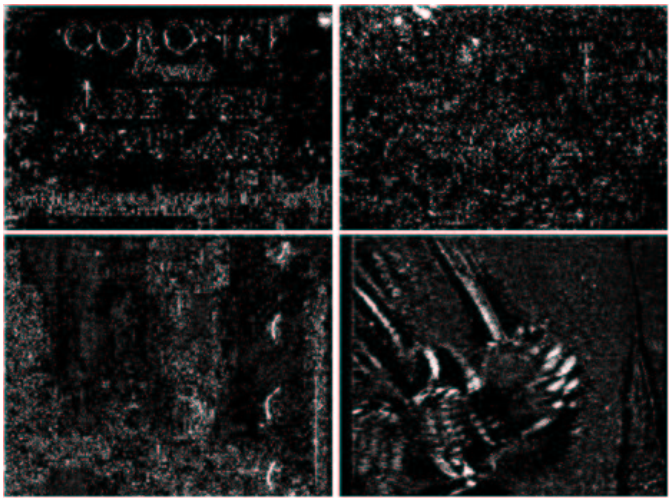


Figure 9: Confidence of dirt extracted for the images in Fig 1.

Using p_n, a confidence image can be further defined as $Conf_n(i,j) = g(p_n(d_n(i,j)))$ where $g(z) = (L-1) * \ln(1+z)/\ln 2$ and L is the number of gray levels in f_n.

Fig 9 provides a visual impression of the estimated confidence for the material shown in Fig 1. From Fig 9 we can see that the results are, on the average, intuitively correct with actual dirt pixels being detected with high confidence and hence appearing bright and vice versa. If a binary dirt detection mask D has already been estimated, a confidence weighted mask D' can be further defined given as

$$D'(i,j) = \begin{cases} Conf_n(i,j) & if \quad D(i,j) \neq 0 \\ 0 & otherwise \end{cases} \qquad (27)$$

Essentially D' is a greylevel image whose non-zero values correspond to dirt particles with an associated confidence value. For a given confidence value $q \in [0,1]$, a final dirt mask D_c can be obtained as

$$D_c(i,j) = \begin{cases} 1 & if \quad D'(i,j) \geq (L-1)q \\ 0 & otherwise \end{cases} \qquad (28)$$

It should be obvious that (27) and (28) do not affect correct detection rates but can be used as a mechanism for controlling false alarms.

4.3 Experimental results

In our experiments, we adaptively combine ROD and SSMF as described above. Fig 10 shows the results obtained using parameters settings same as above with $\lambda = 2$ and $N = 32$. Fig 10 can be compared with detection results achieved using spatial filtering (Fig 3-5) and motion-based filtering (Fig 7-8). We can see that the adaptive scheme is more accurate with fewer false alarms, especially in relation to Figs 1(b) and 1(d). For Figs 1(a) and 1(c), there are still a few residual false alarms, which will be further eliminated using confidence weighting.

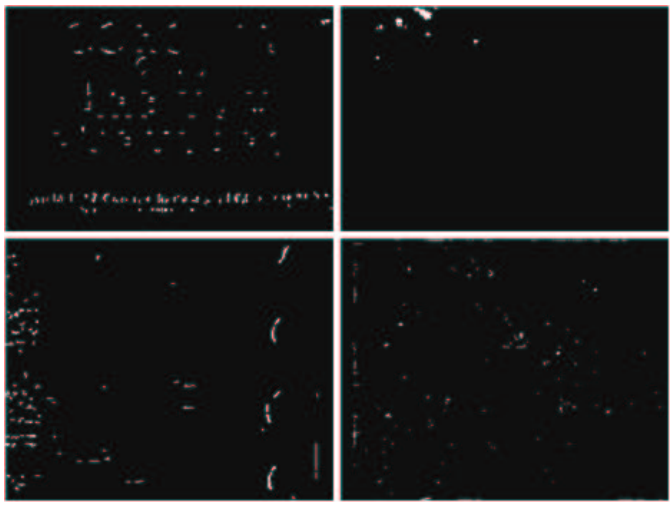


Figure 10: Dirt detected by adaptive combination of ROD and SSMF for the image of Fig 1.

This is shown in Fig 11 which was computed using the confidence images in Fig 9. Comparing Fig 1, Fig 10 and Fig 11, we can confirm that adaptive dirt detection coupled with confidence weighting provides a reliable solution offering a good balance between a high correct detection rate and a low false alarm rate.

It is worth noting that residual false alarms in Fig 11, have low confidence values associated with them and hence can be easily eliminated. For a given confidence level at $q = 0.7$, Fig 12 shows a binary dirt mask of dirt computed using Eq. (28). We can see that most false alarms have been successfully eliminated.

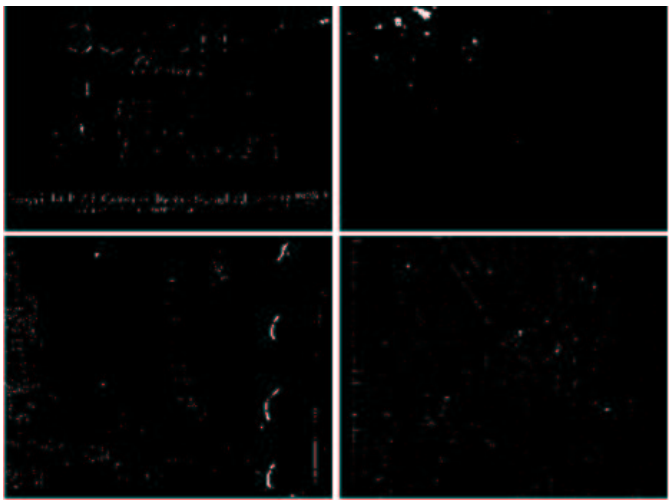


Figure 11: Dirt detection after confidence weighting.

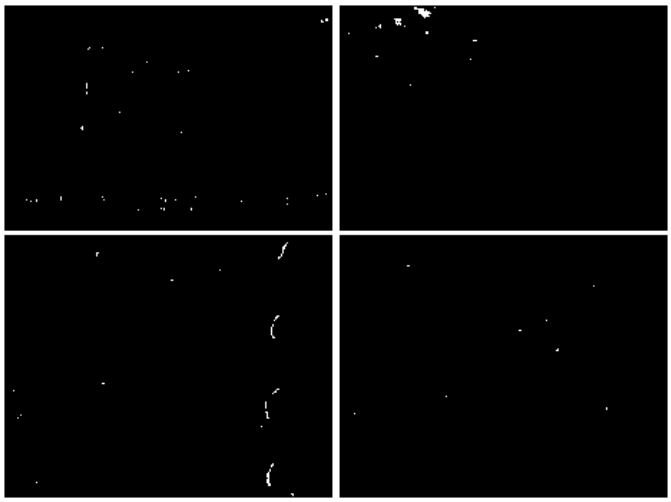


Figure 12: Binary mask of dirt at a confidence level of 70%.

5 Quantitative evaluations and discussion

5.1 Criteria for quantitative evaluation

In Hamid *et al* [6], a corresponding ideal (dirt-free) reference sequence is required in order to carry out an evaluation using the mean absolute error (MAE) criterion. In practice such a reference is unavailable. In our experiments, we employ manually derived ground truth of dirt for quantitative evaluations.

Based on manually generated ground truth, a quantitative performance assessment was carried out using three key criteria namely correct detection rate R_c, false alarm rate R_f and missed detection rate R_m. If D_g is a ground truth dirt mask and D_x is a dirt detection mask obtained from any given method, these criteria are defined as follows:

$$R_c = \frac{Count(D_g \otimes D_x)}{Count(D_g)} \tag{29}$$

$$R_f = \frac{Count(D_x \otimes \overline{D}_g)}{Count(D_g)} \tag{30}$$

$$R_m = \frac{Count(D_g \otimes \overline{D}_x)}{Count(D_g)} = 1 - R_c \tag{31}$$

$$\overline{D}_y(i,j) = \begin{cases} 1 & if \quad D_y(i,j) = 0 \\ 0 & otherwise \end{cases} \tag{32}$$

where $y = g, x$ in D_y, $Count$ is a function counting the non-zero elements in a mask and operator $\otimes$ is the logical AND defined as follows:

$$(D_x \otimes D_y)(i,j) = \begin{cases} 1 & if \quad D_x(i,j) * D_y(i,j) \neq 0 \\ 0 & otherwise \end{cases} \tag{33}$$

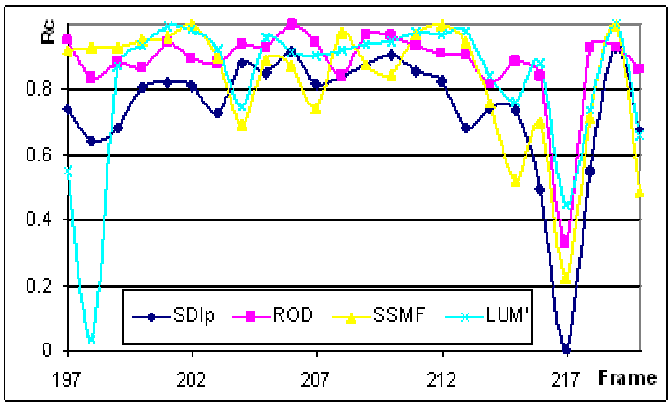
5.2 Results and discussions

Broadcast resolution (720x576 pixels) sequence "Pennine Way" which contains fast motion and textured background was used in this set of experiments. SSMF and LUM' were combined adaptively with SDIp and ROD, respectively.

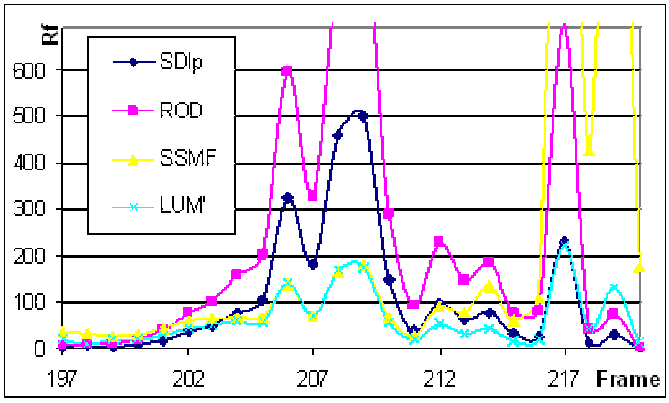
Using Eqs (29) to (31), we can determine $0 \le R_c, R_m \le 1$ and $0 \le R_f \le N_0 / N_g - 1$, where N_0 and N_g are the number of pixels in the input image and number of dirt pixels in D_g, respectively. If $D_g = 200$, we have $R_f \le 2015$ for the "Pennine Way" sequence! In fact we have found a maximum R_f value of 4000 using SSMF, which means too many false alarms have been detected despite the fact that $R_c = 1$.

Fig 13 shows dirt detection performance using SDIp, ROD, SSMF and LUM'. The parameters for these four methods are $t_1 = 15$, $t_2 = 25, t_3 = 45, t_s = 15$ and $r = 3$.

In Fig 13 we can see that the correct detection rate of ROD is higher than SDIp, but the false alarm rate is also higher (a maximum value of nearly 700). Regarding for spatial filtering, LUM' is better than SSMF except for the first two frames and it has a lower false alarm rate on average. The highest value of the false alarm rate (in excess of 1481) occurs for SSMF in frame 217 while the lowest value for SDIp is 0.81 occurring in frame 220. Furthermore, the correct rate varies from 0 (frame 217 using SDIp) to 1 (frame 106 and 219 using ROD and SSMF/ LUM', respectively).



(a) Correct detection rate as a function of frame number



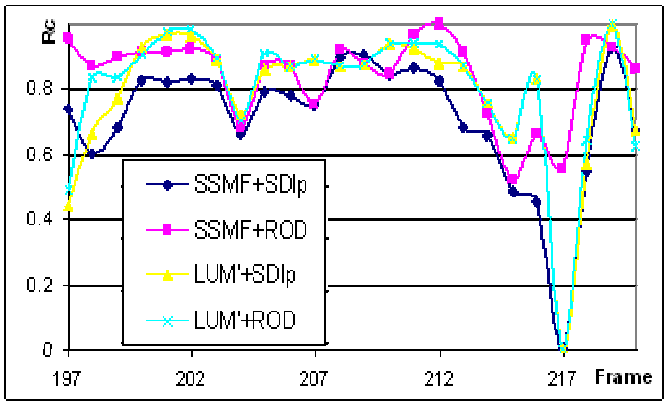
(b) False alarm rate as a function of frame number

Figure 13: Dirt detection performance for "Pennine Way" using SDIp, ROD, SSMF and LUM'.

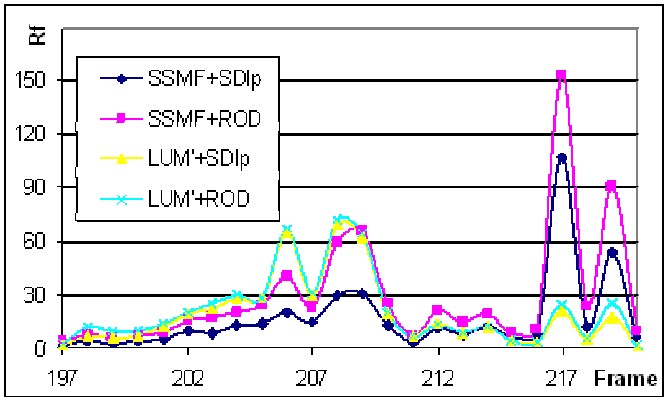
The unstable performance of these methods in Fig 13 means apparently that that they are all very sensitive to image content even within the same sequence. We further test four combination schemes coupling each of SDIp and ROD with each of SSMF and LUM'. For simplicity we keep parameter, $\lambda = 1.11$ constant while the block size used is 32×32. Results are shown in Fig. 14 .

Combining SSMF with ROD offers the best overall performance in terms of correct detection rate. However, for individual frames, this rate may be comparable to those obtained using a stand-alone approach. On the other hand, false alarm performance has improved significantly.

To further reduce the false alarm rate, we use confidence weighting. The results are shown in Fig 15 and Fig 16. In Fig 15 we can observe a substantial false alarm rate in frames 217 and 219. Inspection of Fig 16 confirms a substantial improvement in that respect. As a result, the combination of SSMF with ROD seems the best overall solution for accurate and robust dirt detection offering simultaneously high correct detection rates and low false alarm rates. On the other hand, the combination of SSMF and SDIp generates the lowest false alarm rate but, at the same time, the lowest correct detection rate.



(a) Correct detection rate as a function of frame number



(b) False alarm rate as a function of frame number

Figure 14: Dirt detection performance comparison for "Pennine Way" using combination schemes.

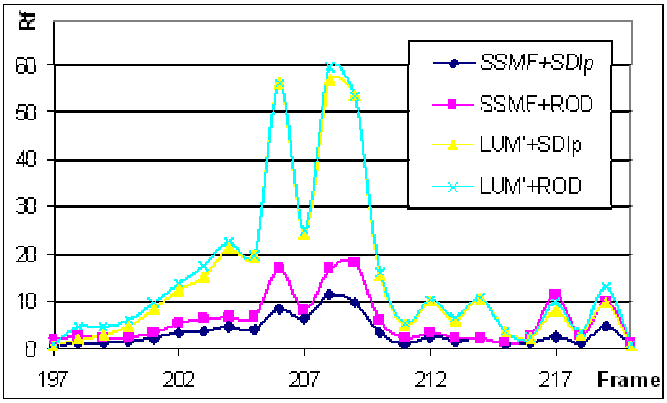


Figure 15: False alarm rate as a function of frame number for combination schemes using confidence weighting for "Pennine Way".

As we can see in Fig 16 confidence weighting can successfully reduce the false alarm rate compared to conventional stand-alone methods.

There are two important design parameters in the proposed adaptive approach, namely block size and parameter, λ whose influence we investigate experimentally employing the best performing method (namely SSMF combined with ROD and using confidence weighting). The results are shown in Fig 17, from which we can see that similar correct detection rates are obtained with block sizes of 16, 32, 48 and 64. Regarding false alarms, it seems that smaller windows cause fewer false alarms. A good tradeoff between correct detection and false alarms can be obtained using block sizes 32 and 48. On average, these results are quite comparable confirming that our method is not sensitive to the choice of block size.

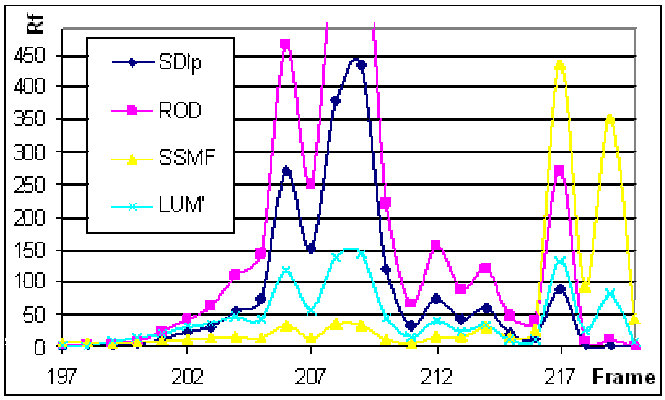
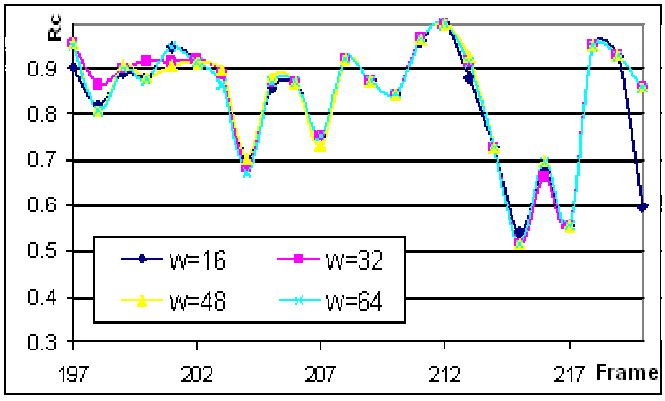
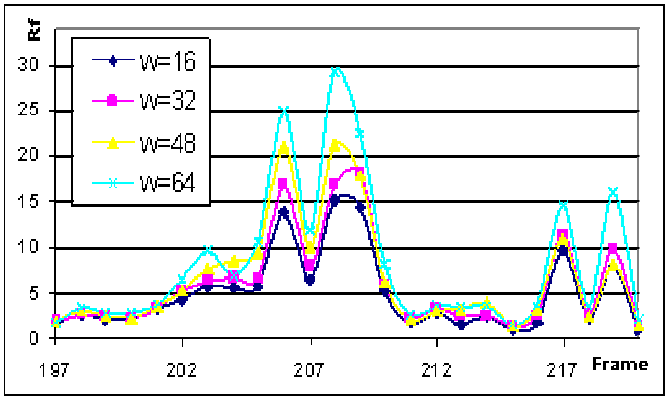


Figure 16: False alarm rate as a function of frame number for stand-alone schemes using confidence weighting for "Pennine Way".



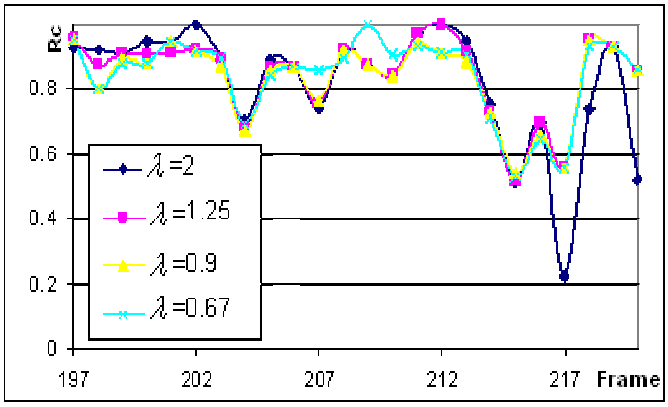
(a) Correct detection rate as a function of frame number



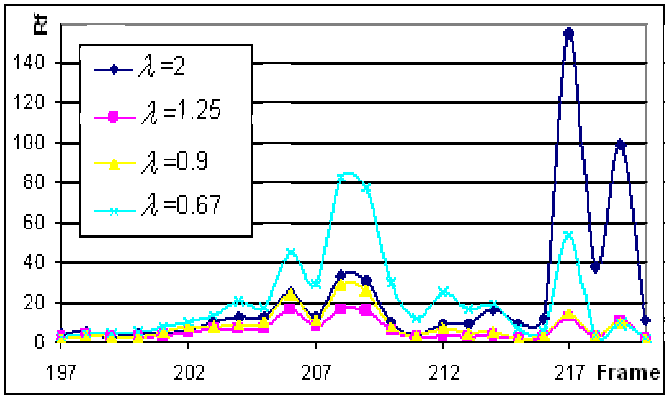
(b) False alarm rate as a function of frame number

Figure 17: Performance comparison using different block sizes.

Regarding parameter λ, Fig 18 shows results obtained using $\lambda = 2.0, 1.25, 0.9$ and 0.67. It seems that very high or very low λ will produce more false alarms and worse correct detection rate while performance is not particularly affected by the parameter taking values in the range between 0.67 and 1.25.



(a) Correct detection rate as a function of frame number



(b) False alarm rate as a function of frame number

Figure 18: Performance comparison using different values of parameter λ.

6 Conclusions

We have presented an adaptive spatio-temporal approach for dirt detection in archive film sequences. The proposed strategy is based on a combination of SSMF and ROD and moreover incorporates a measure of confidence weighting which further improves false alarm rate performance. Overall our results have shown that our scheme offers a practical solution to the dirt detection problem and compares favourably with conventional approaches.

Acknowledgements

This work was carried out in the framework of project PrestoSpace, supported by the European Commission under Framework Programme 6. The authors would like to thank the research staff at Snell & Wilcox for valuable discussions and for providing some of the test data.

References

[1] B. Alp, P. Haavisto, T. Jarske, K. Oistamo, Y. Neuvo. "Median-based algorithms for image sequence processing", *in Proc. SPIE*, **1360**, pp. 122-134, (1990).

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

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

[4] N. C. Gallagher, G. L. Wise. "A theoretical analysis of the properties of median filters", *IEEE Trans. Acoust., Speech, Signal Proc.*, **29**, pp. 1136-1141, (1981).

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

[6] M.S. Hamid, N.R. Harvey, S. Marshall. "Genetic algorithm optimization of multidimensional grayscale soft morphological filters with applications in film archive restoration", *IEEE Trans. Cir. Sys. for Video Tech.*, **13(5)**, pp. 406-416, (2003).

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

[8] A.C. Kokaram. "Motion picture restoration", Berlin, Germany: Springer-Verlag, (1998).

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

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

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

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

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

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