Non-motion-compensated region-based dirt detection for film archive restoration

Jinchang Ren
University of Surrey
Center for Vision, Speech and Signal Processing
Guildford, GU2 7XH, United Kingdom
and
Northwestern Polytechnic University
School of Computers
Xi'an, 710072, China
E-mail: J.Ren@surrey.ac.uk

Theodore Vlachos
University of Surrey
Center for Vision, Speech and Signal Processing
Guildford, GU2 7XH, United Kingdom

Abstract. A novel non-motion-compensated method is proposed for dirt detection in archived film sequences. A confidence measurement extracted from raw differences between current frame and each of the previous and next frames is used to exploit the temporally impulsive nature of dirt impairments. Further evolution of the confidence signal enables the minimization of false alarms and the fine-tuning of detector sensitivity. After morphological filtering and consistency checks candidate regions of dirt emerge, enabling the computation of binary detection masks. Our experiments show that our method compares favorably with extended spike detection index (SDIp), rank order detector (ROD), and lower-upper-middle (LUM) approaches and provides efficient and robust detection performance for a wide range of archived film material. © *2006 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.2338592]

Subject terms: Film dirt detection; archive restoration; video signal processing; video enhancement.

Paper 050907R received Nov. 19, 2005; revised manuscript received Jan. 20, 2006; accepted for publication Jan. 30, 2006; published online Aug. 28, 2006.

1 Introduction

Film archive restoration is a key enabling technology toward facilitating access to historical footages. By reducing the perceptual impact of archive-related impairments, restoration with improved quality can meet viewers' aesthetic expectations and enrich their viewing experiences. Furthermore, the suppression of these impairments has vital implications on the efficiency of video coding algorithms used in the television and multimedia context such as MPEG-2. Consequently, it has recently attracted a lot of interest and several high-profile collaborative projects have received European Union (EU) 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 toward 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.[1–4] In this paper, our focus is on the detection of impairments occasionally referred to as dirt. These are among the most commonly encountered artifacts and hence their successful detection is a priority issue in any archive restoration system.[1] 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 that are attached to the film or localized abrasions that occurred during storage or when the film passed through various transport mechanisms.[1,3] Figure 1 gives some examples of dirt which are shown bounded by rectangles.

In this paper, we propose an efficient and robust method for film dirt detection. Compared with other methods in the literature, a unique feature of our algorithm is that a confidence measure is derived for dirt detection, which is also attached to the detected results. This is an invaluable feature toward both automatic and operator-assisted dirt concealment as it enables a variable degree of treatment according to preference. A second attractive feature of our scheme is that we detect dirt as regions consisting of hard core and soft boundaries rather than as isolated pixels. Thus, we can fine-tune detection performance using variable confidence levels. Finally, we do not employ motion estimation and motion-compensated prediction, which reduces the complexity of our scheme considerably and makes it a good candidate for fast implementation.

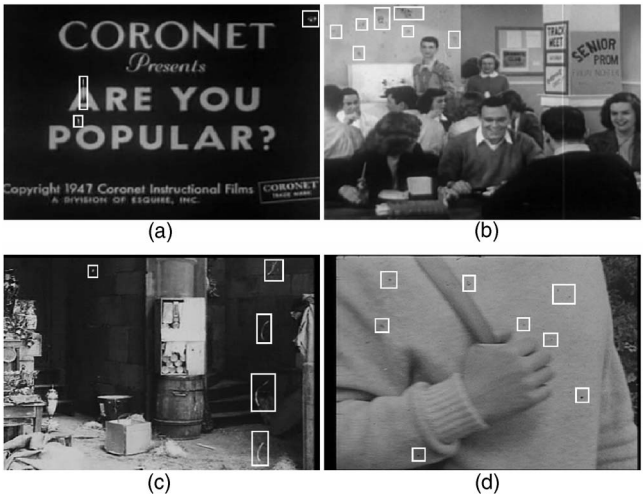

Fig. 1 Examples of dirt: (a) static text, (b) low-motion natural scene, (c) moderate-motion natural scene, and (d) fast-motion natural scene.

2 Related Work

Owing to its temporal characteristics, interframe processing has proved a useful tool toward detection and concealment of dirt. Consequently, a common approach is to apply temporal median filtering on the current frame and its two motion-compensated neighbors.[2,5–8] In Schallauer et al.,[2] a pixel is filtered as dirt if both its absolute differences between the 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 Kokaram,[5] the so-called "spike detection index" (SDI) is proposed. This is also based on the identification of high absolute differences between the current frame and two compensated images. The extended SDI method, SDIp, additionally requires sign consensus of the two differences above. Nadenau and Mitra[6] proposed the rank order detector (ROD), in which a total of 7 pixels from three consecutive frames are compared against three thresholds. Gangal et al.[7] extended ROD to five frames to improve accuracy in heavily corrupted images or occluded blotches. Poosmalen et al.[8] presented a simplified version of ROD, in which only one threshold was utilized, but complex postprocessing is required to remove false alarms.

Since motion-based approaches require a high degree of complexity and can be unreliable when motion estimation fails, many spatial filtering techniques have also been proposed as alternatives.[3,9] Most of these methods employ median filtering, as it can preserve more edges than linear filters.[10] Subsequently, dirt detection is based on the identification of high difference values between the current frame and the filter output. Hardie and Boncelet[9] proposed LUM (lower-upper-middle) filters, which utilized two parameters for adjustable smoothing and sharpening of images. Compared with soft morphological filters (SMFs), Hamid et al.[3] pointed out that LUM failed to restore fast-moving objects in image sequences. However, SMFs seem impractical for most applications as they require sufficient representative dirt samples for training to obtain optimal filters.

3 Extracting Confidence Image

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 differencing. Then, we define d_n as

$$d_n=\begin{cases}\dfrac{2d_{n-}d_{n+}}{|d_{n-}|+|d_{n+}|} & \text{if } d_{n-}d_{n+}>0\\ 0 & \text{otherwise.}\end{cases} \tag{1}$$

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

$$p_n(m)=\lambda_0^{-1}\int_{m_0}^{m}p_0(x)\mathrm{d}x, \tag{2}$$

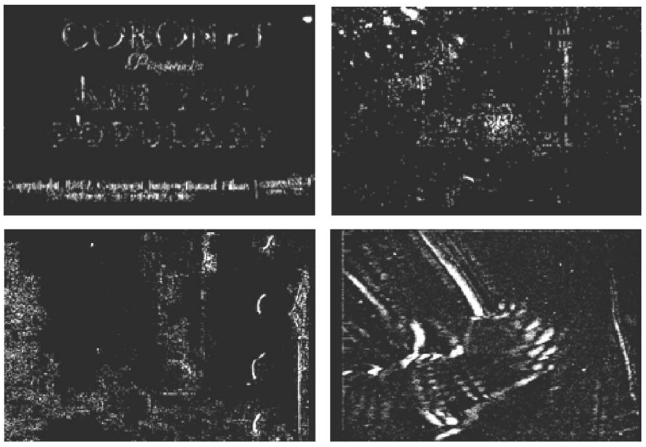

Fig. 2 Four confidence images extracted from the four images in Fig. 1.

$$\lambda_0+\int_0^{m_0-1}p_0(x)\mathrm{d}x=1, \tag{3}$$

where p_0 is the original probability density function (pdf) of d_n, λ_0 is used to normalize p_n within [0, 1], and m_0 is used to control the removal of static background.

Note that $m_0=0$ in Eq. (2) amounts to histogram equalization of d_n. Nevertheless, a static background in the three consecutive frames may force most pixel values in d_n near zero, thus straightforward histogram equalization is not useful in this context. Let μ, γ, and σ be the mean, median, and variance of the intensity distribution of d; in this case, we always have $\gamma\leqslant\mu$. Then m_0 is determined by

$$m_0=\frac{\mu+\gamma}{2}+\sigma. \tag{4}$$

With p_n, a new image, Conf, can be obtained in Eq. (5) representing confidence of dirt. Since $p_n(m)\in[0,1]$, we have $\mathrm{Conf}_n(i,j)\in[0,255]$.

$$\mathrm{Conf}_n(i,j)=255*p_n[d_n(i,j)]. \tag{5}$$

Figure 2 shows four confidence images extracted for the images in Fig. 1. Comparing Figs. 1 and 2, we can find that the dirt areas in the original images have been successfully located in the confidence images. However, there are also many false alarms in the confidence images. How to detect dirt from these confidence images and to remove false alarms is given in Sec. 4.

4 Region-Based Dirt Detection

With the extracted confidence information, a simple way for determining dirt pixels is to threshold the confidence image. An adjustable confidence level c_q can be used to obtain a binary mask of detected result, mask_n, which is given by

$$\mathrm{mask}_n(i,j,c_q)=\begin{cases}1 & \text{if } \mathrm{Conf}_n(i,j)\geqslant 255c_q\\ 0 & \text{otherwise.}\end{cases} \tag{6}$$

Fig. 3 Another group of confidence images extracted from the four images in Fig. 1.

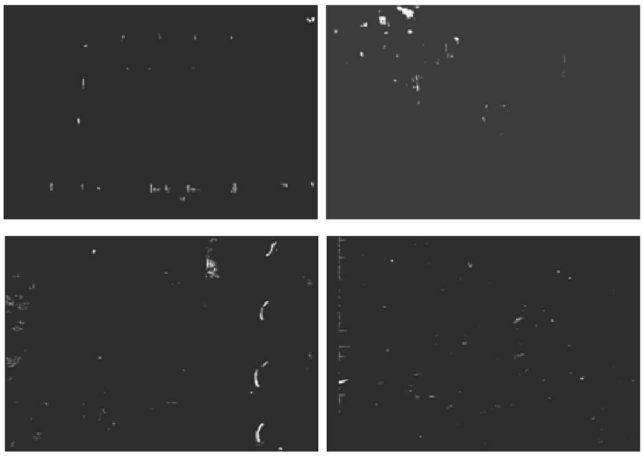

Fig. 4 Gray-scale images of dirt detected from the four images in Fig. 1.

If there is no fast motion in the sequences, we can set a higher c_q for dirt detection. Otherwise, a lower c_q is necessary to extract dirt of lower intensity from massive false alarms of higher intensity (see the bottom-right image in Fig. 3).

4.1 *Filtering*

The difference between dirt and false alarms is that the prior is motion-independent, but the latter is mainly caused by (camera and/or object) motion. Therefore, we extract another confidence image, Conf'_n, from f_{n-2}, f_n, and f_{n+2}, and assume that regions of dirt in the two confidence images should be more consistent than those of false alarms. For the confidence images in Fig. 2, the corresponding Conf'_n images are shown in Fig. 3.

For a predefined confidence level of c_q, we can extract two binary masks mask_n and mask'_n from the confidence images Conf_n and Conf'_n, respectively. For robustness, morphological close operation is applied to the gray-scale confidence image before extraction of the binary mask using Eq. (6). Then, all the connected regions in each mask are labeled.

Let R and R' be two regions in mask_n and mask'_n, respectively, and $R \cap R' \neq \text{NULL}$. Then, two new regions R_2 and R'_2, as the corresponding region of R and R', are defined in mask'_n and mask_n as

$$\begin{cases} R_2 = R \cap \text{mask}'_n \\ R'_2 = R' \cap \text{mask}_n \end{cases}. \tag{7}$$

Let s, s', s_2, and s'_2 be the sizes of R, R', R_2, and R'_2, respectively. Then, R is a region of dirt if we have

$$\begin{cases} s_{\max} \geqslant s \geqslant s_2 \\ s' \leqslant \lambda_s s'_2 \end{cases}. \tag{8}$$

In Eq. (8), $\lambda_s > 1$ is a constant, and $s_{\max}$ is the maximum area of dirt regions. In our experiments, we set $\lambda_s = 1.6$ and $s_{\max} = 220$. After checking all the regions in mask_n, only those regions that satisfy Eq. (8) will remain as detected dirt.

4.2 *Confidence Attachment*

Let B_n be the new binary mask obtained after removal of false alarms. Confidence information is attached to B_n to obtain a gray-scale dirt image. We avoid using Conf_n because false alarms are contained in d_n (see Sec. 3). Instead more accurate confidence images are extracted as follows.

First, a new measurement of dirt D_n is obtained as

$$D_n(i,j) = \begin{cases} d_n(i,j) & \text{if } B_n(i,j) \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Then, D_n is employed to generate a new confidence image C_n in the same way as to extract Conf_n from d_n. To obtain a binary mask of dirt, finally we threshold C_n using various confidence levels, as in Eq. (6).

Figure 4 shows the gray-scale images of detected dirt from the four images in Fig. 1. From Figs. 1(a)–1(d), the predefined confidence levels (to extract B_n) for the images in Fig. 4 are 0.9, 0.85, 0.8, and 0.7, respectively. From Fig. 4 we can see that our method produces quite accurate results in which most of the false alarms are removed from the confidence images.

5 Evaluation and Discussions

We compare our method with three other established algorithms, SDIp (Ref. 5), ROD (Ref. 6), and LUM (Ref. 9). For motion compensation in SDIp and ROD, the well-known Black-Anandon optical flow algorithm of subpixel accuracy is utilized.[11]

5.1 *Methods in Comparison*

Let f_n, C_{n-}, and C_{n+} be the current frame and the two motion-compensated frame neighbors. We then define D_{n-} and D_{n+} as the differences between each of these two images and the current frame, i.e., $D_{n-} = C_{n-} - f_n$ and $D_{n+} = C_{n+} - f_n$.

Using the preceding notation we subsequently consider established approaches such as SDIp (Ref. 5) and ROD (Ref. 6). In Ref. 5, dirt pixels are defined as

$$D_{\mathrm{SDIp}} = \begin{cases} 1 & \text{if } |D_{n-}(i,j)| > t_s,\ |D_{n+}(i,j)| > t_s,\ D_{n-}D_{n+} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

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

$$D_{\mathrm{ROD}} = \begin{cases} 1 & \text{if } e_1 > t_1 \text{ or } e_2 > t_2 \text{ or } e_3 > t_3 \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

In LUM, two parameters k and l are introduced for smoothing and sharpening, respectively. Typically it holds that $1 \leqslant k \leqslant l \leqslant N_0 = (N+1)/2$, and N is the total number of pixels in a given window. Let the central pixel in the original current frame and filtered image be denoted as x' and y', respectively, and $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(N)}$ be the rank-ordered pixels within the window. Then, the filtered output is defined as

$$y' = \begin{cases} x^L & \text{if } x' \leqslant (x^L + x^U)/2 \\ x^U & \text{otherwise,} \end{cases} \tag{12}$$

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

$$\begin{cases} x^L = \mathrm{median}\{x_{(k)}, x', x_{(l)}\} \\ x^U = \mathrm{median}\{x_{(N-k+1)}, x', x_{(N-l+1)}\}. \end{cases} \tag{13}$$

In Arce,[12] LUM was further extended to a $3 \times 3 \times 3$ spatial-temporal window, where we have $N=27$ and $k \leqslant 14$. The output of the filter is given by

$$y' = \mathrm{median}[x_{(k)}, x', x_{(N-k+1)}]. \tag{14}$$

Finally, pixels of dirt are determined by

$$D_{\mathrm{LUM}} = \begin{cases} 1 & \text{if } \mathrm{abs}(y' - x') \geqslant t_p \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

5.2 *Visual Assessment*

Using SDIp ($t_s=25$), ROD ($t_1=15, t_2=25, t_3=45$), and LUM ($t_p=10$), the detected results of dirt from the four images in Fig. 1 are shown in Fig. 5, and t_s, t_1, t_2, t_3 and t_p are the relevant thresholds in the corresponding algorithms. From Fig. 5 we can see that SDIp generates fewer false alarms than ROD and LUM, but ROD produces more accurate detection than SDIp. Moreover, LUM is the worst as it has the most false alarms but the least accuracy.

In SDIp, ROD, and LUM, one or more thresholds are required for dirt detection. Normally, larger thresholds will lead to fewer false alarms but less accuracy, thus suitable thresholds are essential in dirt detection. Unfortunately, these thresholds cannot be determined automatically. In Ref. 5, t_s is chosen to be within 7 to 45. In ROD, the important first threshold t_1 changes from 1 to 38! Our algorithm only uses thresholds for parameters related to false alarm filtering and, more importantly, these thresholds are substantially more stable than the above with no performance loss.

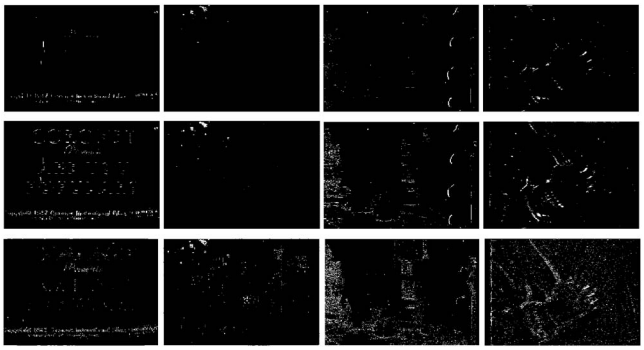

Fig. 5 Detected dirt from the four images in Fig. 1 using SDIp (top row), ROD (middle row), and LUM (bottom column).

Furthermore, dirt detected in our method appears as gray-scale images with attached confidence, which contains regions of dirt with higher intensities in the center and lower intensities around. This is similar to the results from Ref. 3 in which dirt is detected using a model-based structure with a hard core and soft boundaries. Therefore, decreasing confidence level will produce more detail of detected dirt in the binary masks, as it enables more soft boundaries to be detected.

5.3 *Quantitative Assessment*

We obtain a quantitative performance assessment using manually derived ground truth maps and two key criteria, namely, correct detection rate R_c and false alarm rate R_f. Let D_g be a binary mask representing the ground truth of dirt, and D_x is the binary mask of detected dirt from either SDIp, ROD, LUM, or our method. Then, we can define R_c and R_f as

$$R_c = \frac{\mathrm{Count}(D_g \otimes D_x)}{\mathrm{Count}(D_g)}, \tag{16}$$

$$R_f = \frac{\mathrm{Count}(D_x \otimes \bar{D}_g)}{\mathrm{Count}(D_g)}, \tag{17}$$

where Count is a function counting the nonzero elements in a mask and operator $\otimes$ is the logical AND between two binary masks.

For the four sequences in Fig. 1, the quantitative comparisons of detected results from different methods are given in Table 1. Since manually derived ground truth is more reliable in identifying a whole region of dirt in spite of the soft boundaries, a relevant lower confidence level is necessary for more accuracy in our method. Therefore, dirt results detected under confidence levels of 0.7 and 0.5 are taken for comparisons. As LUM produces worst results, only SDIp and ROD are compared.

Table 1 Quantitative performance comparison of SDIp, ROD, and our method at confidence level of C_q=0.7 and C_q=0.5, respectively.

	Items							
	Average Correct Detection Rate				Average False Alarm Rate			
Data	SDIp	ROD	C_q=0.7	C_q=0.5	SDIp	ROD	C_q=0.7	C_q=0.5
Seq. 1(a)	0.371	0.659	0.578	0.643	12.55	108.49	20.64	33.53
Seq. 1(b)	0.376	0.682	0.553	0.655	0.62	30.21	3.79	7.09
Seq. 1(c)	0.512	0.802	0.795	0.888	16.60	106.8	7.96	13.66
Seq. 1(d)	0.312	0.453	0.746	0.932	22.88	46.11	4.58	6.86

From Table 1 we can see (1) on the average our method yields the most accurate detection rate and the lowest false alarm rate; (2) a lower confidence level in our algorithm will lead to a few more false alarms, but the accurate rate can be improved apparently; (3) lower detection rate is obtained for interpolated video sources, i.e., sequences of Figs. 1(a) and 1(b) resampled from 15 to 30 frames/s, this is because dirt may become no longer an ideal one-frame event after interpolation; and (4) due to failure of motion estimation, SDIp and ROD may occasionally yield very poor detection results of quite low correct rate and high false alarm rate (see Fig. 6). Figure 6 illustrates an example of frame-based performance comparison when motion estimation fails, in which the broadcast resolution (720 $\times$576) "Pennie Way" sequence from Fig. 1(d) is taken, which contains fast motion and textured background.

5.4 *Computational Efficiency*

Furthermore, we also compared the complexity of the four algorithms. For "Pennie Way," SDIp and ROD require about 53.93 s to process one frame, in which about 99% of the time (53.38 s) is spent for bidirectional motion estimation. Although there are motion vectors available in many compressed video streams, say MPEG-2, they are too sparse to be used in such a context. Therefore, Black-Anandan algorithm is employed for motion estimation to generate dense motion vectors (for each pixel) of subpixel accuracy. On the same machine, LUM takes about 8.5 s for each frame, and our method requires only 0.25 s, which is about 36 times and 216 times faster than LUM and SDIp (or ROD), respectively.

6 Conclusions

We presented a novel method for film dirt detection. One advantage of our approach in relation to techniques operating in the spatial domain is the lack of any requirement to specify the shape and size of a filter. Another relative advantage is the capability to detect dirt as a region that is consistent with actual manifestations of dirt. A key advantage is the attached confidence in the detected result, which enables fine-tuning of detector sensitivity. Finally, relative to motion estimation methods, our scheme has the much lower computational complexity.

Acknowledgments

This work forms part of the PrestoSpace project, supported by the European Commission FP6-507336. The authors would like to thank the research staff at Snell & Wilcox PLC, Hampshire, United Kingdom, for valuable discussions and for providing some of the test data.

References

1. A. C. Kokaram, "On missing data treatment for degraded video and film archives: a survey and a new Bayesian approach," *IEEE Trans.*

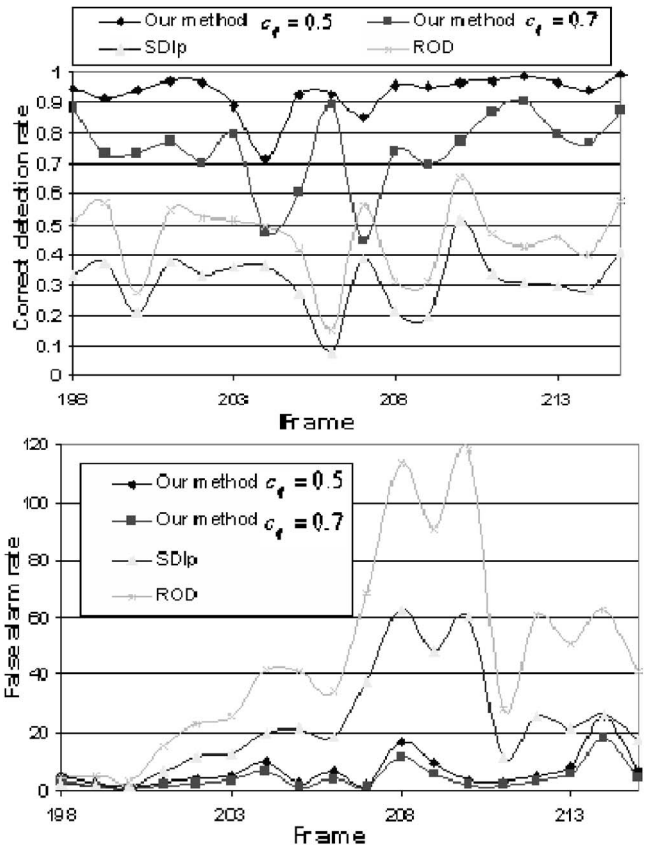

Fig. 6 Frame-based quantitative comparisons of our algorithm with SDIp and ROD in terms of correct detection rate (top) and false alarm rate (bottom).

Image Process. **13**(3), 397–415 (2004).
2. P. Schallauer, A. Pinz, and W. Haas, "Automatic restoration algorithms for 35 mm film," *J. Comput. Vis. Res.* **1**(3), 59–85 (1999).
3. M. S. Hamid, N. R. Harvey and S. Marshall, "Genetic algorithm optimization of multidimensional greyscale soft morphological filters with applications in film archive restoration," *IEEE Trans. Circuits Syst. Video Technol.* **13**(5), 406–416 (2003).
4. A. C. Kokaram, R. Morris, W. J. Fitzgerald, and P. J. W. Rayner, "Detection of missing data in image sequences," *IEEE Trans. Image Process.* **40**(11), 1496–1508 (1995).
5. A. C. Kokaram, *Motion Picture Restoration*, Springer-Verlag, Berlin (1998).
6. M. J. Nadenau and S. K. Mitra, "Blotch and scratch detection in image sequences based on rank ordered differences," in *Proc. 5th Int. Workshop Time-Varying Image Processing*, pp. 27–35 (1996).
7. A. Gangal, T. Kayikcioglu, and B. Dizdarogle, "An improved motion-compensated restoration method for damaged color motion picture films," *Signal Process. Image Commun.* **19**, 353–368 (2004).
8. P. M. B. Poosmalen, J. Biemond, and R. L. Lagendijk, "Restoration and storage of film and video archive material," *J. Signal Process. Multimed.* **1**(3), 167–191 (1999).
9. R. Hardie and C. Boncelet, "LUM filters: a class of rank-order-based filters for smoothing and sharpening," *IEEE Trans. Signal Process.* **41**, 1061–1076 (1993).
10. N. C. Gallagher and G. L. Wise, "A theoretical analysis of the properties of median filters," *IEEE Trans. Acoust., Speech, Signal Process.* **29**, 1136–1141 (1981).
11. M. J. Black and P. Anandan, "The robust estimation of multiple motions: parametric and piecewise-smooth flow fields," *Comput. Vis. Image Underst.* **63**(1), 75–104 (1996).
12. G. Arce, "Multistage order statistic filters for image sequence processing," *IEEE Trans. Signal Process.* **39**, 1146–1163 (1991).

Jinchang Ren received his BEng, MEng, and PhD degrees in 1992, 1997, and 2000, respectively, all from Northwestern Polytechnic University (NWPU), Xi'an, China, where in 1992 he became a teaching assitant with the Department of Computer Science and Engineering. He became a lecturer in 1997 and an associate professor in 2001. He is currently a research staff member with the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, United Kingdom. His research interests are video processing, including content-based analysis and retrieval, motion detection and tracking, and archive restoration, and image and video segmentation and recognition.

Theodore Vlachos received his DiplIng degree from the University of Patras, Greece, in 1985, his MSc degree from the University of Maryland, College Park, in 1986, and his PhD degree from Imperial College, London, in 1993. He joined the BBC Research and Development Department, Kingswood Warren, United Kingdom in 1993, where he led projects on video compression and archive restoration. During that time, he became an accredited member of the United Kingdom delegation at ISO-MPEG, contributing to the work of MPEG-4. His work on archive restoration and image stabilization gained international recognition and resulted in patents, prizes, and highly successful commercial products. In 1997, he become a lecturer in multimedia signal processing with the Centre for Vision, Speech and Signal Processing, University of Surrey. He is a chartered electrical engineer and a member of the IEE and the Technical Chamber of Greece.