Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 1735–1746
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Fusion of complementary detectors for improving blotch detection in digitized films

Sorin Tilie [a,b,*], Isabelle Bloch [b], Louis Laborelli [a]

[a] *Département Recherche et Expérimentation, Institut National de l'Audiovisuel, 4 av. de l'Europe, 94366 Bry sur Marne, France*
[b] *GET-ENST, TSI CNRS UMR 5141, 46 rue Barrault, 75013 Paris, France*

Received 16 November 2006; received in revised form 22 April 2007
Available online 18 May 2007

Communicated by R. Davies

Abstract

This paper proposes a novel method based on Dempster–Shafer (belief function) theory for the detection of blotches in digitized archive film sequences. The detection scheme relies on the fusion of two uncorrelated blotch detectors, one working in the spatial domain and the other one in the temporal domain. The imprecision and uncertainty of both detectors have been modeled using belief function theory, and their combination improves the decision, by taking into account the ignorance and the conflict between detectors. Quantitative evaluation using real blotches ground truth shows that this combination scheme improves the global performance, and compares favorably with two classical blotch detectors.

PACS: 07.05.Mh; 07.05.Pj

Keywords: Digital film restoration; Spatio-temporal filtering; Dempster–Shafer data fusion

1. Introduction

The last century has provided a large amount of audiovisual documents, which are endangered because of the smooth, but irreversible degradation of their support. The main challenge for the traditional stakeholders is to achieve the migration to digital formats and the long-term preservation of these digitized documents (http://www.prestospace.org/).

Digital restoration is a key step, because of the increasing need in image quality required by the new digital high definition (HD) broadcast formats (HDDVD, HDTV and HD cinema). However, digital restoration has to deal with several (sometimes contradictory) constraints: improve the visual quality of the restoration to deal with high-resolution images (HD, 2k or 4k), decrease the costs, which are closely related to the automation of the process (operator costs), to the hardware and its evolvability, and finally to the speed of the restoration process. These constraints should be fulfilled by the introduction of software solutions, working on standard low-cost PCs, and by the research of new, fast, adaptive, and high-level algorithms.

Film is the oldest and the most fragile moving picture media. Film impairments are related to the storage conditions (moisture, vinegar syndrome, dye fading), to improper handling (scratches, dust, dirt), and to poorly maintained equipment (scratches, unsteadiness).

In this paper we focus on the detection of the most frequent defects, which are dirt and sparkle, grouped under the term of "blotches". Dirt and sparkles are impulsive

* Corresponding author. Address: Département Recherche et Expérimentation, Institut National de l'Audiovisuel, 4 av. de l'Europe, 94366 Bry sur Marne, France. Tel.: +33 0149833292; fax: +33 0149832929.
E-mail addresses: stilie@ina.fr (S. Tilie), Isabelle.Bloch@enst.fr (I. Bloch), llaborelli@ina.fr (L. Laborelli).

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2007.05.005

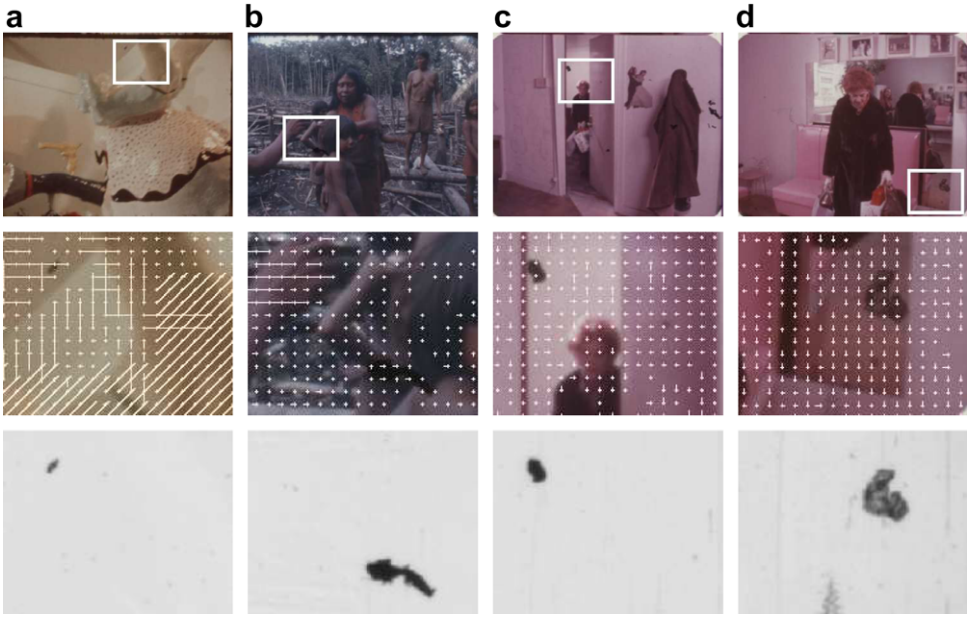

Fig. 1. Examples of blotched 16 mm color film images. Original images showing examples of blotches in the white boxes (top row), zoomed parts (4×) scanned in RGB with motion vectors superimposed (middle) and in infrared (bottom). (a) Fast camera motion natural scene, (b) low motion but textured natural scene, (c and d) low motion natural scenes.

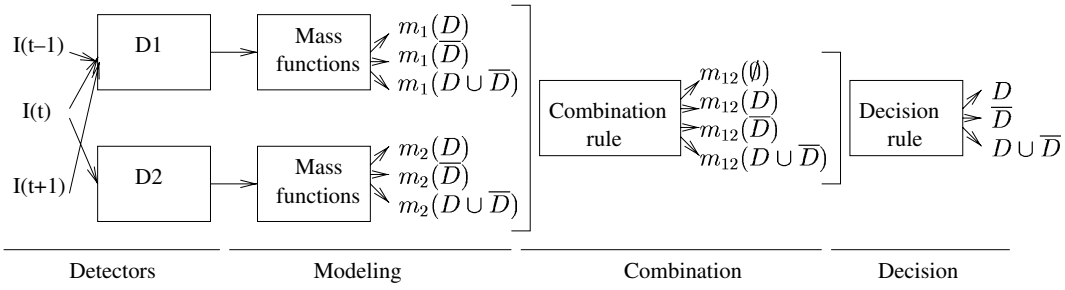

Fig. 2. Main steps of the fusion process: blotch detection by the temporal blotch detector $D1$ (taking three consecutive images $I(t-1)$, $I(t)$ and $I(t+1)$) and by the spatial blotch detector $D2$, numerical output modeling by mass functions m (corresponding to hypotheses "defect" (denoted by D), "no-defect" (denoted by $\overline{D}$) and "ignorance" (denoted by $D \cup \overline{D}$)), masses combination, and decision.

(single frame) defects. Dirt can be seen as opaque or semi-transparent clusters with random size, shape and position, caused by dust and dirt stuck on the film, while sparkle are white clusters, caused by the local abrasion of film gelatin.

Fig. 1 illustrates some examples of blotches on 16 mm color film, scanned both in RGB and in infrared. The first image shows an opaque blotch and global motion, the second and third ones show opaque blotches and local motion, while the fourth image shows opaque and semi-transparent blotches with low motion.

In this paper, we propose a method based on the Dempster–Shafer (DS), or belief function, fusion framework (Shafer, 1976; Smets, 1990) for optimizing the combination of two blotch detectors. It extends the preliminary approach presented in (Tilie et al., 2006). In particular, the mass function estimation in now automatized, and a more extensive performance evaluation is performed. The schema in Fig. 2 shows the processing flow: imprecision and uncertainty of each detector modeling followed by their combination in order to take the best decision given this partial information. The fusion also provides a risk index, allowing a variable degree of processing.

The remainder of this paper is organized as follows. Section 2 gives an overview of the previous blotch detection methods, while Section 3 describes two blotch detectors that we combine in Section 4. Section 5 shows experimental results, and conclusions are presented in Section 6.

2. Previous works

Dirt and sparkle detection often relies on the assumption of smoothness of motion in the sequence, and on temporal and/or spatial inconsistency of defects. Usually, the image is filtered in the temporal and/or spatial domain, and the defects are detected as the thresholded difference between the original frame and the filtered one.

Temporal detection methods have been used early in the problem of dirt concealment (Storey, 1985). Basically, these methods are based on the computation of the displaced frame difference (DFD), which is the difference between two consecutive motion compensated (MC) frames. Kokaram introduced the spike detection index (SDIa) (Kokaram and Rayner, 1993) which computes the minimum between backward and forward DFD, and SDIp that reduces the number of false alarms, by introducing an additional constraint that requires the signs of both DFDs to be identical before a blotch can be detected. In a similar way, Schallauer et al. (1999) introduced the constraint that the absolute differences of previous and next MC frames have to be below a threshold.

In spite of good global performances, these methods have high computational costs (because of motion estimation), and produce false alarms when motion estimation fails, i.e. for so-called "pathological motion" (Rares et al., 2001) (occlusions, uncovering, intermittent motion, erratic motion, motion blur, large displacements, transparency).

Spatial filtering methods assume spatial inconsistency of defects. Median filters have been extensively used because of their ability to eliminate outliers while preserving edges. Nieminen et al. (1987) proposed a multi-level median filter (MMF), which was implemented as the median of the outputs of several median filters with different topologies. Hardie et al. introduced the Lower–Upper–Middle (LUM) filter (Hardie and Boncelet, 1993), in which the output is assigned to the upper or to the lower local median values computed on an upper and a lower intervals defined by two parameters. Buisson et al. (1997) used a top hat morphological filter, because of its ability to detect specific patterns, such as dust and hair.

However, these spatial filters provide false alarms on sharp and textured regions, i.e. when image spatial patterns look like defects patterns, and fail to detect blotches exceeding the filter size.

Spatio-temporal methods extend spatial filtering to the temporal domain, often using MC frames. Arce (1991) introduced a multi-stage order statistic filter (MOS) as an extension of the min/max MMF filter to three (non-MC) frames. In the same manner, Alp et al. (1990) presented the ML3D filter, which uses median operations and provides, according to Kokaram (1993), better impulse noise rejection than that proposed by Arce. Kokaram (1993) improved ML3D filter (ML3Dex) by using three MC frames. Nadenau and Mitra (1996) presented the rank order detector (ROD), which compares the rank ordered differences between the current pixel and six neighbors (from the previous and next image) against three thresholds. Roosmalen (1999) proposed a simplified ROD (SROD) by using a single threshold for the maximum distance between the current pixel and the minimum or maximum of the neighbors. Gangal et al. (2004) improved ROD performances by robust motion estimation based on seven motion vectors computation, and extending ROD to five MC frames. More recently, they presented a fuzzy pre-filtering, followed by a less expensive motion estimation based on three motion vectors computation, more robust with respect to impulsive blotches, that improves blotch reconstruction (Gangal and Dizdaroglu, 2006). Unfortunately, no evaluation of the improvement in blotch detection is reported in the article. Buisson et al. (1997) presented a hybrid detector, based on the combination of SDIa with a spatial morphological filter, while Decenciere Ferrandiere and Serra (1997) extended the class of morphological area filters to the spatio-temporal domain.

An original adaptive filter was recently presented by Hamid et al. (2003). A soft morphological filter (SMF) works on three non-MC images, and its size and shape parameters are learned using a genetic algorithm, supervised by both artificially corrupted and uncorrupted sequences. This filter gives less false alarms than LUM and ML3Dex for fast moving objects, but the learning step is very slow, and the filter parameters should be adapted to each new sequence.

Probabilistic methods have also been proposed, using a Bayesian framework and Markovian models for the detection of spatial and temporal discontinuities on MC frames (Bornard, 2002; Kokaram, 1998; Morris, 1995). These methods perform well in real situations, but have a high computational cost, which become untractable when the neighborhood order exceeds first or second order (Bornard, 2002). Bornard (2002) extended the Markov random fields MRF model to five MC frames, and introduced an interpretation step (based on spatio-temporal redundancy of false alarms), that reduced the number of false alarms within pathological motion areas, but also reduced the number of good detections.

Spatio-temporal methods achieve better performance than spatial or temporal methods alone. However, the computational load is higher, and often false alarms due to "pathological motion" persist, which can be seen as a consequence of a sub-optimal combination.

A novel spatio-temporal method for dirt detection in color films has been introduced in (Ren and Vlachos, 2007). This method combines confidence extracted from the color of previous/next no-motion compensated frames and local spatial region growing, to reduce the number of false alarms. This approach has a low complexity because of the lack of motion compensation, but for the same reasons it seems to be applicable to sequences with limited local motions only.

A better modeling of the combination of these detectors should improve the quality of detection and decrease the computational time. Temporal detection should be used to leave the ambiguity between dirt clusters and objects of a similar spatial structure, while spatial detection should be used to confirm temporal detections, or to replace these ones when motion estimation fails.

The fusion method proposed in this paper exploits these ideas.

3. Blotch detectors

Two blotch detectors have been chosen, providing redundant and complementary detections. Belief function theory only assumes detectors cognitive independence. Cognitive independence is weaker than the statistical independence, and only assumes that the detectors have no information about each other (Shafer, 1976). The cognitive independence of these detectors can be assumed, since detectors act in two different domains (one detects temporal discontinuities, while the second detects spatial discontinuities) and have no information about each other. Then it makes sense to use a non-idempotent operator, as the classical conjunctive belief function combination, to benefit from its reinforcement effect.

Both detectors are designed to work on intensity images, as many archive films are black and white. For color films, intensity is computed as the Y channel in the YCbCr color space.

3.1. Temporal detector

The temporal detection is achieved by a SROD detector (Roosmalen, 1999), working on three frames. For a given frame to be processed, the previous and next frames can be motion compensated, using the well known phase correlation algorithm (PhC) (Thomas, 1987). The SROD detector labels a pixel in the current frame as "blotched", if its intensity value is an outlier of the intensity distribution of its neighbors taken in the MC previous and next images (Fig. 3).

The "outlierness" of a pixel in the current image is computed as the difference between the minimum of the distribution and the pixel value for a black outlier, or as the difference between the pixel value and the maximum of the distribution for a white outlier. Let $r_{\min} = \min_{k=1\ldots N}(r_k)$ and $r_{\max} = \max_{k=1\ldots N}(r_k)$, where N is the number of neighbors of a pixel x and r_k the intensity value of the kth neighbor (from the previous/next image). The detector output at pixel x is then defined as:

$$SROD(x) = \max\{0, r_{\min} - I(x), I(x) - r_{\max}\}, \quad (1)$$

where $I(x)$ is the value of the current pixel x.

This detector has one free parameter S, which is the size of the square neighborhood in the previous and next images $\left(\text{where } S = \sqrt{\frac{N}{2}}\right)$. This parameter should be larger than the residual motion between two consecutive images, but small enough to avoid the loss of sensibility around edges.

Fig. 4 shows examples of opaque and semi-transparent blotches well detected by our temporal detector, and an example (in the first image) of false alarm, caused by two strong (backward and forward) DFDs, due to the motion estimator inability to correctly estimate (backward and forward) fast motion.

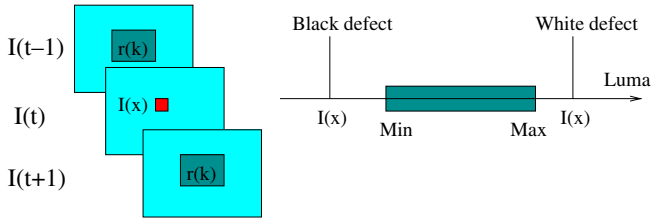

Fig. 3. Principle of SROD: SROD is the difference of intensity between a pixel in the current image and the minimum/maximum of the distribution of its neighbors in the previous and the next images.

3.2. Spatial detector

In order to decrease the number of false alarms provided by the previous detector, we introduce a second detector, acting in the spatial domain. As blotches are often local extrema with sharp edges in the images, we assume that they can be detected by a morphological filter.

Initially, a top hat morphological filter using a non-flat large square structuring element with a tunable profile (Buisson et al., 1997) has been implemented. It provides good results for the detection of sharp dust, but its profile is not well adapted to the detection of large and semi-transparent blotches, which are frequent on 16 mm film (see Fig. 1). As the performances of these morphological filters are limited by the great variety of blotch shapes and sizes, we have relaxed the constraints on the blotch shape to a single constraint on the blotch area.

Area black (white) top hat morphological filters are the residuals of the morphological area opening (closing) operators introduced by Vincent (1993). In contrast with classical mathematical morphology, area operators are insensitive to the actual shape of detected regions. Area opening (closing) operator cuts (fills) a peak (gap or valley) until the area of the cuts (filled) peak (valley) exceeds a given area value. Image structures which do not satisfy this increasing criterion on area are left unchanged.

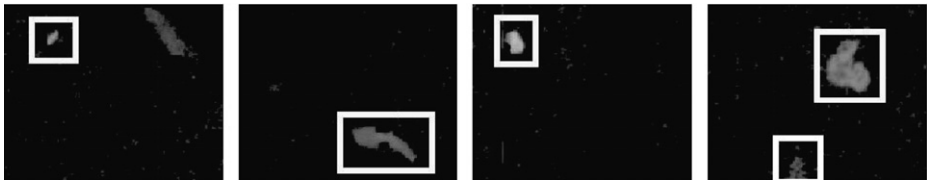

Fig. 4. Temporal detections of blotches, corresponding to the images in Fig. 1. White boxes show examples of semi-transparent and opaque blotches well detected, while remaining detections are caused by pathological motion.

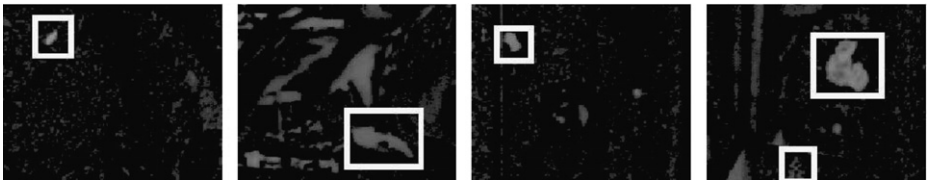

Fig. 5. Spatial detection of blotches, corresponding to images in Fig. 1. White boxes show examples of semi-transparent and opaque well-detected blotches. Remaining detections are false alarms caused by objects having spatial patterns close to blotches ones.

Morphological area opening and closing have been efficiently implemented using Tarjan's union-find algorithm (Wilkinson and Roerdink, 2000), with a reduced computational cost.

This detector has one free parameter λ, which corresponds to the maximal area allowed for the detected regions. Its value can be considered as a prior knowledge, which depends on the film size (16 mm or 35 mm), on the scanning resolution (SD or HD), and on the blotches size which is related with their type (dust, dirt, sparkle). Given our test footage (16 mm scanned at standard resolution), we have fixed this parameter to a large value (500 pixels), to allow the detection of defects of various sizes, especially large semi-transparent blotches and elongated structures (hair).

Fig. 5 shows examples of blotches well detected by our spatial detector, examples of semi-transparent blotches partially detected because of their low contrast, and examples of false alarms, due to spatial image patterns having a size close to blotches size.

However, we can see that good detections confirm the ones obtained with the temporal detector, while false alarms are globally different in the two detections results. This way, redundant detections reinforce the detection, while false alarms are detected as a "conflict" between detectors.

4. Data fusion

In order to take advantage of the redundancy and complementarity of both detectors, we combine their outputs in the framework of the belief function (or Dempster–Shafer, DS) theory, which easily handles the concepts of uncertainty, ignorance and imprecision of data.

4.1. Belief structures

We assume that each pixel can be explained by two mutually exclusive and exhaustive hypotheses: "defect" denoted D or "non-defect" denoted $\overline{D}$. The frame of discernment is denoted by Θ:

$$\Theta = \{D, \overline{D}\}. \tag{2}$$

The set of subsets of Θ is then:

$$2^{\Theta} = \{\emptyset, \{D\}, \{\overline{D}\}, \Theta\}. \tag{3}$$

The DS theory introduces the concept of disjunction (or compound hypothesis), to allow the representation of the ignorance of a source of information (i.e. if the source cannot distinguish between hypotheses), by assigning the corresponding belief to the union of these hypotheses. A special case is the total ignorance (in the case of missing data), which can be modeled by assigning a non-null belief to the union of all hypotheses (Θ). In our two-class application, the only compound hypothesis, denoted $D \cup \overline{D}$, is equal to Θ, and accounts for the detector inability to classify a pixel as D or $\overline{D}$ near the boundary between classes.

In the DS theory, each hypothesis is associated with a mass in the interval $[0, 1]$. Masses corresponding to hypotheses in 2^{Θ} should fulfill the normalization constraint:

$$\begin{cases} m(\emptyset) & = 0, \\ \sum\limits_{A \in 2^{\Theta}} m(A) & = 1. \end{cases} \tag{4}$$

In our application, for each pixel, each detector provides a numerical value in the interval $I = [0, 255]$. Hence each value of $m(A)$ is actually a function defined on I, i.e. depending on the strength of the detection result. A function on the image space can also be derived, by assigning to each pixel the value of $m(A)$ derived for the value of the detection at this pixel. We will denote indifferently $m(A)(x)$ or $m(A)(f(x))$ where $f(x)$ denotes the value in I at pixel x.

The definition of the mass functions is not straightforward, because no general method exists. The main problem comes from the estimation of masses of compound hypotheses. Methods based on probabilistic schemes have been introduced by Appriou (1993) or Dromigny-Badin (1998). They assume that the conditional distributions of probabilities $p(f(x)|A)$ (where $f(x)$ is the detector output for the pixel x) are known. In our application the ground truth of defects is available, and the conditional probabilities $p(f(x)|D)$ and $p(f(x)|\overline{D})$ can be estimated (on a "learning" set) by the normalized histograms of the detectors outputs, with respect to the corresponding binary ground truth masks (see Section 5).

Appriou's models assume the knowledge of a set of discounting parameters specific to each detector and to each hypothesis. As detectors reliabilities are difficult to estimate in our application, we chose Dromigny's method.

In this method, compound hypotheses have been defined using a significance criterion for the conditional probabilities (a threshold, arbitrary set to a low value (10^{-6})). Several situations can occur, as described next.

First, if none of the conditional probabilities is significant, the full mass is assigned to Θ:

$$m(\Theta) = 1. \tag{5}$$

Second, if only one conditional probability $p(f(x)|A)$ is significant, the mass on A is given by the value of the conditional probability:

$$m(A)(f(x)) = p(f(x)|A). \tag{6}$$

Finally, if several conditional probabilities are significant, disjunctions of hypotheses are taken into account. In our two-class application, masses are given by:

$$\begin{cases} m(\{D\})(f(x)) = \max(0, p(f(x)|D) - p(f(x)|\overline{D})), \\ m(\{\overline{D}\})(f(x)) = \max(0, p(f(x)|\overline{D}) - p(f(x)|D)), \\ m(\{D \cup \overline{D}\})(f(x)) = \min(p(f(x)|\overline{D}), p(f(x)|D)). \end{cases} \tag{7}$$

For each case, masses are then normalized, to achieve the normalization constraint (Eq. (4)).

Fig. 6 shows the empirical distributions of probabilities, conditionally to hypotheses D ($p(f(x)|D)$) and $\overline{D}$ ($p(f(x)|\overline{D})$), corresponding to the numerical outputs of (motion compensated) temporal and spatial detectors. These conditional distributions have been obtained from normalized occurrence frequencies of the detector output conditionally to ground truth masks on a "learning" sequence.

In order to eliminate the histogram noise, conditional probabilities have been smoothed by a non-linear function. Several functions have been tested, and we found that a sum of two decreasing exponentials gave the best fitting:

$$\hat{p}(f(x)|A) = C_1 \exp(-\lambda_1 x) + C_2 \exp(-\lambda_2 x). \tag{8}$$

The linear parameters C_1, C_2 and non-linear parameters λ_1, λ_2 have been estimated by the Nelder-Mead simplex optimization method (Lagarias et al., 1998) for $A = D$ and $A = \overline{D}$ and for the two detectors. Fig. 7 shows the conditional probabilities fitted by this function.

We can see that high $p(f(x)|\overline{D})$ values correspond to lower detection values, while $p(f(x)|D)$ values are distributed along higher detection values. However, some significant $p(f(x)|D)$ values occur for lower detection values, because some defects have been missed or weakly detected. Also, the distribution of $p(f(x)|\overline{D})$ of the temporal detector is sharper than the corresponding distribution of the spatial detector, because the second detector generates more false alarms than the first one.

Fig. 8 shows the set of mass functions computed for each detector, according to the previous method.

We can see that the ignorance is maximum between the D and $\overline{D}$ hypotheses, and decreases to zero with the increasing temporal detector output values, while it decreases

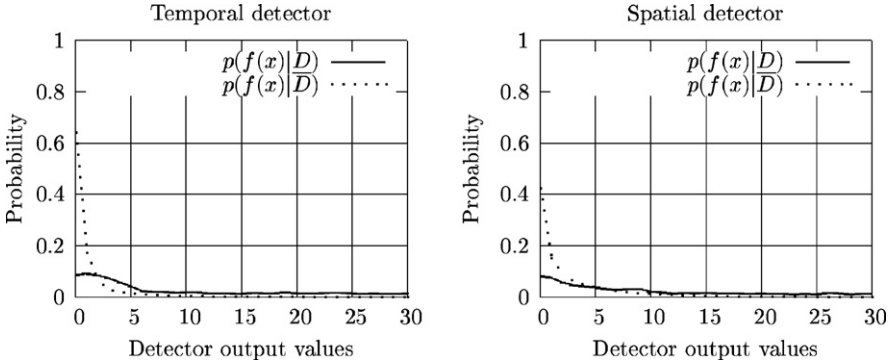

Fig. 6. Distributions of the probabilities conditionally to hypotheses D and $\overline{D}$. Left: temporal detector; right: spatial detector. For visualization purposes, only the values in the interval $[0,30]$ have been shown.

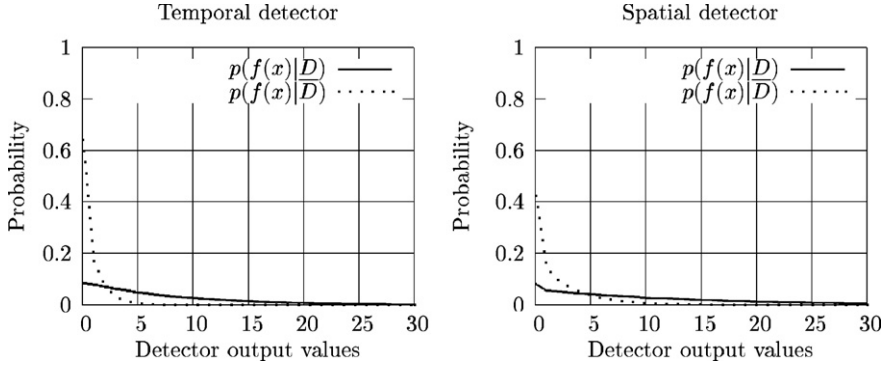

Fig. 7. Conditional distributions of probabilities, fitted by a non-linear function. Left: temporal detector; right: spatial detector.

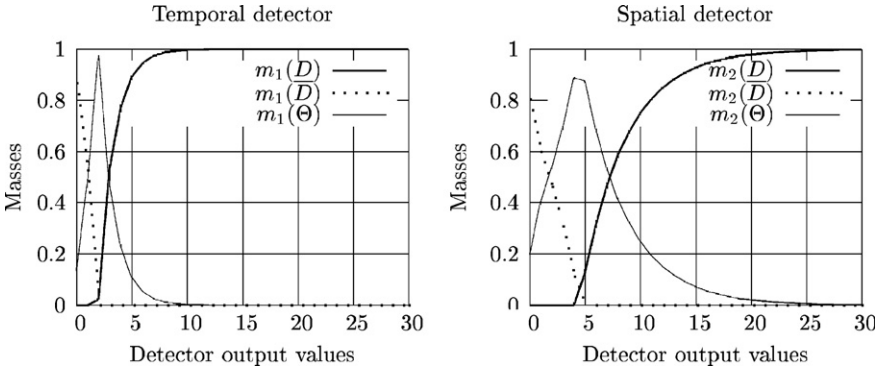

Fig. 8. Mass functions associated to the temporal detector (left) and to the spatial detector (right), function of the detectors output values.

more slowly for the spatial detector, because this detector is less accurate.

Given the two sets of mass functions, we can now associate to each pixel two sets of masses (corresponding to temporal and spatial detections).

4.2. DS combination

Previously defined masses can now be combined pixel-wise, using the unnormalized combination rule (Shafer, 1976; Smets, 1990):

$$m_{12}(A) = \sum_{B \cap C = A} m_1(B) m_2(C). \tag{9}$$

The intersection table corresponding to this combination rule is illustrated in Table 1. This table also shows the principle of the DS fusion: the transfer of mass from the combined hypotheses to simple hypotheses, in order to reduce the ignorance.

This combination rule assigns a mass to the empty set ($m_{12}(\emptyset)$), which is interpreted as a measure of the conflict between detectors. This conflicting mass can be reassigned to other hypotheses, and the employed method is highly dependent on the problem modeling.

Several solutions have been proposed for the conflict management (Lefevre et al., 2002). If we consider that detectors are perfectly reliable, we can use Smets' rule of combination (Eq. (9)), i.e. without re-assigning the conflicting mass (open world assumption), or the Dempster's rule which corresponds to Eq. (9) normalized by $1 - m(\emptyset)$ (closed world assumption). However, for highly conflicting sources, normalization hides the conflict, which is a problem in some applications.

If we consider that detectors are not reliable, we can discount the detectors provided that we can estimate their reliability (Shafer, 1976), or use a disjunctive rule of combination (Lefevre et al., 2002; Smets, 1990), which has the drawback to increase the ignorance.

In our application, we consider that the frame of discernment (see Section 4.1) was correctly modeled, but sometimes detectors fail (closed world assumption). For example, the first detector may fail for complex motions, while the second detector fails for texture patterns which are similar to blotch patterns.

Thus, masses should be discounted by the reliability of each detector. Unfortunately, the reliability of the temporal and spatial detectors cannot be estimated as one global factor.

For these reasons, we chose Dubois and Prade (1988) rule of combination, where the conflicting mass of two subsets B and C is assigned to their union (compound hypothesis) $B \cup C$:

$$m_D(B \cup C) = m_{12}(B \cup C) + \sum_{B \cap C = \emptyset} m_1(B) m_2(C). \tag{10}$$

We can notice that in our two-class model, Dubois's rule simplifies to Yager (1987) rule, which assigns the conflicting mass to the whole set Θ (total ignorance):

$$m_Y(\Theta) = m_{12}(\Theta) + m_{12}(\emptyset). \tag{11}$$

Yager's rule of combination merges the total ignorance and the conflict, which can raise some semantics problems in some applications. However, this allows defining a single index of ignorance, as long as we consider that the conflict is due to the fact that we ignore which detector is unreliable.

Table 1
Intersections between the sets of hypotheses of temporal and spatial detectors

	D	$\overline{D}$	Θ
D	D	$\emptyset$	D
$\overline{D}$	$\emptyset$	$\overline{D}$	$\overline{D}$
Θ	D	$\overline{D}$	Θ

4.3. Decision rule

The decision is usually based on the belief and plausibility functions, which are the minimum and the maximum of uncertainty of a given hypothesis A:

$$\begin{cases} Bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B), \\ Pls(A) = \sum_{B \cap A \neq \emptyset} m(B). \end{cases} \tag{12}$$

Fig. 9. Ignorance ($m(\Theta)$) corresponding to images in Fig. 1.

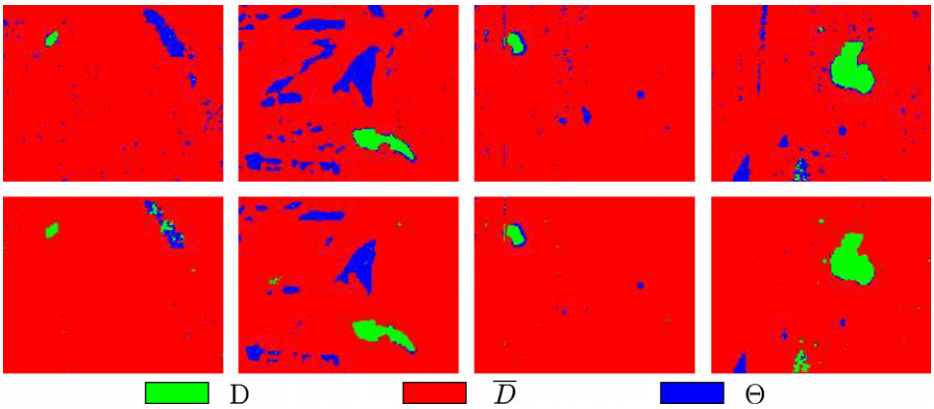

Fig. 10. Decisions corresponding to images in Fig. 1. (top) Low risk decisions ($Th = 0.1$), (bottom) high risk decision ($Th = 0.6$).

The length of the belief interval $[Bel(A), Pls(A)]$ can been interpreted as the imprecision about the uncertainty value, and in our two singleton hypotheses case it is equal to $m(\Theta)$.

The decision rule we have chosen selects the singleton hypothesis which corresponds to the maximum of belief if the total ignorance is lower than a given threshold Th; otherwise, the decision is Θ (total ignorance):

$$\begin{cases} D & \text{if } Bel(D) > Bel(\overline{D}) \quad \text{and} \quad m(\Theta) \leqslant Th, \\ \overline{D} & \text{if } Bel(\overline{D}) > Bel(D) \quad \text{and} \quad m(\Theta) \leqslant Th, \\ \Theta & \text{else.} \end{cases} \tag{13}$$

Fig. 9 shows the ignorance attached to each pixel. The ignorance results from the ignorance masses attached to each detector in the modeling step, and from the mass of conflict computed in the combination step. The ignorance mass has been used to provide a "risk of decision" index, which weights the decision in favor of simple "defect" or "no-defect" hypotheses, against the "ignorance" hypothesis.

Fig. 10 shows the decisions corresponding to this risk index thresholded with $Th = 0.1$ and $Th = 0.6$. We can see that redundant detections of semi-transparent blotches have been classified as "defects", and that (non-redundant) false alarms provided by spatial or temporal detectors have been rejected in the "ignorance" class, because of the conflict between detectors.

Fig. 10 also shows the trade-off between good detections and false alarms resulting from this risk index thresholding: low values ($Th = 0.1$) ensure "sure" (low risk) decisions, but many pixels are rejected in the "ignorance" class, which decreases the number of false alarms and the number of good detections. A higher value ($Th = 0.6$) enforces the decision in favor of D and $\overline{D}$ hypotheses, which increases the number of good detections, but also increases the number of false alarms.

Finally, to remove the isolated pixels, a morphological opening with a disk of radius 1 as structuring element has been performed on the binary mask corresponding to decision D. Resulting masks will be evaluated in the next section, using the ground truth extracted from infrared images.

5. Experimental results

The performance of our blotch detection method has been compared with the performance of two classical blotch detection algorithms, using the ground truth of four test image sequences.

5.1. Test sequences

Test sequences have been scanned by a flatbed scanner for film with an additional infrared channel, and resampled to comply with video standard broadcast resolution (720×576). Archive materials are 16 mm color films from national French TV archives, and have been chosen to show challenging situations illustrating the diversity of defects (blotches, sparkles, scratches), as well as the diversity of video content (fast local and global motion, occlusion, small random patterns) (cf. Fig. 1).

The first sequence, called "Art" (83 images) shows an indoor scene, with strong camera motion and static background. This sequence is quite difficult because of strong

moving texture in the foreground, and patterns size similar to blotches size. The rate of corruption (estimated as the ratio between the number of corrupted pixels and the total number of pixels in the sequence) is low (0.00027) with respect to other sequences, and defects are medium and small (about 150 pixels) opaque and semi-transparent blotches along with thin structures called "hair".

The second sequence, "Cigarette" (269 images) shows an outdoor scene, with small moving textures in the background, and moving persons in the foreground. The difficulty here comes from the fact that blotches are small and hardly visible, because of the dark textured background. The rate of corruption is higher (0.00038) and defects are large–medium (about 400 pixels) opaque and semi-transparent blotches, hair and persistent scratches.

The third and fourth sequences, called "Door" (264 images) and "Surprise" (246 images) have been scanned from the same reel, and show a static indoor scene and a woman performing complex motions. The amount of corruption is higher (0.00048 and 0.00038), and defects are large–medium (more than 400 pixels) opaque and semi-transparent blotches, hair and some persistent scratches.

5.2. Ground truth

In order to perform a quantitative evaluation, the ground truth of defects is required. A traditional approach consists in doing manual segmentation of blotches, which is tedious and difficult to do for a large set of images. Another solution consists in generating artificial blotches of random shape, location, and gray levels, but with a constant gray level value over the blotch area. Unfortunately, this model is unrealistic, as real blotches can be semi-transparent, non-uniform, and have large areas (about 500 pixels at the standard resolution scanning).

A novel interesting and original method to evaluate the quality of blotch detections using infrared scanning has been introduced in the framework of the PrestoSpace project (Tilie et al., 2006; Ren and Vlachos, 2007), and is used here as well. Infrared images provide the location and the transparency of "physical" defects of a film such as blotches, scratches and gelatin abrasion.

The binary ground truth has been generated by thresholding infrared images (cf. Fig. 1). Choosing a threshold value is not straightforward, as it has an influence on the detectors performance evaluation. In this paper, we manually set the threshold value (170 for 8-bit infrared images) to find binary ground truth patterns as close as possible to the human perception of these defects.

However, this method has some limitations. First it only works for color films, as black and white films are not transparent to infrared light. Second, defects which have been revealed in infrared can be "hidden" in the color image by dark areas, and thus become hardly detectable by our method. Moreover, results can be biased by the presence of non-impulsive defects (scratches) which cannot be detected by the temporal blotch detectors.

5.3. Evaluation

A qualitative evaluation is shown in Fig. 11, comparing the binary detections masks (pixels corresponding to decision D) with the binary ground truth masks, in terms of well detected pixels, false alarms, and detections missed.

Opaque and semi-transparent blotches have been generally correctly detected, even in areas with motion. Some detections have been missed, corresponding to low contrasted blotches hidden by dark areas in the image, and to the pixels situated on the blotches smooth edges. However, some false alarms occur on some fast moving textures, showing the limitation of our approach: if small moving patterns are detected by both detectors, false alarms are kept in the fusion result, giving a false detection.

The quantitative performance assessment was achieved using the classical receiver operator characteristic (ROC) curves. For each detector, the ROC curve plots the good detection rates against the false alarm rates, for different values of one free parameter. If we denote by DE the binary mask of detection, GT the binary mask of ground truth and $\overline{GT}$ its complement in the image, good detections and false alarms rates are defined as follows:

$$\begin{cases} P_c = \frac{|DE \cap GT|}{|GT|}, \\ P_f = \frac{|DE \cap \overline{GT}|}{|\overline{GT}|}. \end{cases} \tag{14}$$

We can notice that the classical definition of P_f implies normalization by the number of "sane" pixels; as the number of sane pixels is very important, P_f takes very low values. However, this should not be a problem, as we look for a comparative evaluation.

A comparative evaluation has been carried out, by computing the ROC curves of several detectors on the same test sequences. Our motion compensated fusioned detector (MCFD) has been compared with two classical (motion

Fig. 11. Blotch detections (corresponding to decisions in favor of D hypothesis in Fig. 10) compared with ground truth images.

compensated) blotch detectors SDIp and MRF, described in (Kokaram, 1998). Motion compensation was achieved using the PhC with sub-pixel accuracy algorithm (Thomas, 1987).

For our fusioned blotch detector, we have set the temporal detector parameter to $S = 5$ to account for noise and small errors in motion compensation, and the spatial detector parameter to $\lambda = 500$ to detect all blotches of size less than 500 pixels. This value is an upper bound of the blotches size distribution, and has been empirically chosen to allow the detection of semi-transparent and elongated blotches, that are frequent in our test sequences, scanned at standard resolution from 16 mm footage. In order to compute the operating points, only the "decision threshold" free parameter Th has been varied from 0 to 1 with a step of 0.1. To show the influence of the choice of the learning set on the decision, two sets of mass functions have been computed by the method described in Section 4.1, using 20 images extracted from two different sequences (named "Dance" and "Tierce").

The SDIp "DFD threshold" free parameter ranged from 5 to 40 in steps of 5. The MRF detector has two free parameters "spatial and temporal penalties for discontinuities", ranging from 5 to 10 and from 3 to 30, respectively. For this detector, only the operating points corresponding to the "best" combinations have been shown.

Fig. 12 shows the resulting ROC curves, for the four test sequences. The vertical line measures the ratio of "corrupted" pixels among the "sane" pixels in each sequence, and can be seen as an "index of corruption" for each sequence. We consider that the operating points should be on the left side of this line, as we consider that the average rate of false alarms should be lower than the average rate of corrupted pixels in the sequence.

Given this operating domain, our fusioned algorithm generates less false alarms than the two other classical blotch detection methods. However, for high values of the "risk of decision" parameter ($Th > 0.7$), our fusioned detector provides more false alarms, resulting from the enforcement of the decision in favor of simple hypotheses (D or $\overline{D}$) of pixels with high ignorance resulting from conflicting spatial and temporal detections. To avoid this phenomenon, Th parameter should be limited to reasonable values (less that 0.7 in this example).

We can notice that the operating domain of our detector is limited by the range of the free parameter Th and depends on the choice of the sequence used for mass functions computation. To show the influence of the choice of the learning sequence on the results, we have drawn the ROC curve for two different learning sequences (of 20 images). Fig. 12 shows little difference between the ROC curves, which tends to prove that the choice of the learning sequence has no noticeable impact on the results.

The fusion method relies on the correct modeling of the data. In our application, this is guaranteed by several factors: firstly the behavior of the detectors is taken into account in the design of the mass functions, secondly parameters are learned automatically and we have shown

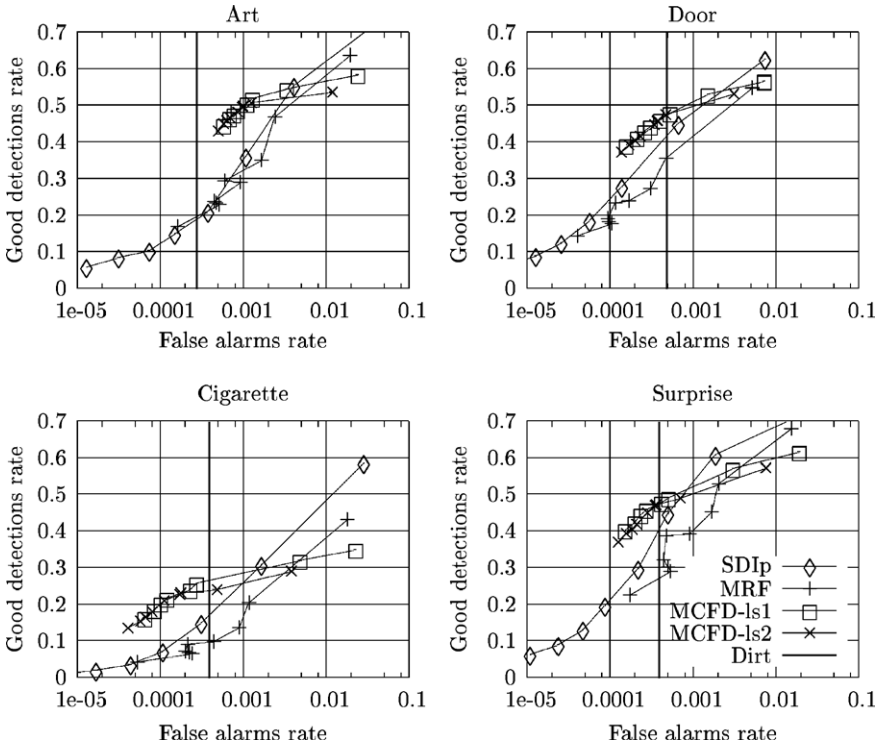

Fig. 12. Average ROC curves corresponding to the SDIp, MRF, MCFD with two learning sets (ls1 and ls2) detectors for the four test sequences. Vertical lines show the average amount of corrupted pixels in each sequence.

that the choice of the learning sequence provides a robust estimation and has no influence on the results, and finally, although different forms could have been chosen instead of the one of Eq. (8), this one fits well the data and has proved to be robust.

The difference of performance between test sequences can also be explained by the content of these sequences. Poor good detection rates in "Cigarette" sequence can be explained by the low visibility of blotches in the images, that makes them difficult to detect. This is rather a limitation of our infrared ground truth, which reveals some blotches hidden by dark areas in the image. Better performance has been achieved for "Art", "Door" and "Surprise" sequences, where blotches were more contrasted in the images. "Art" sequence shows more false alarms, as the large motions have been poorly compensated by the motion estimator.

Finally, our fusioned detector allows the detection of semi-transparent and opaque blotches, and reduces the number of false alarms generated by the temporal or the spatial detector alone. However, it provides false alarms when both temporal and spatial detectors generate false alarms. This is due to the lack of information about the "local" (pixelwise) reliability of spatial and temporal detectors. Local reliability of temporal detectors is a difficult issue, as no algorithm exists in the literature for the detection of pathological motion. In the same way, the estimation of the spatial detector reliability is also a difficult problem, as there are no algorithms to differentiate blotches from similar texture patterns. Reliability estimation is an important issue, and should be explored in further research. Another research issue is the introduction of more spatial coherence in the detections. Actually, spatial coherence of detections relies only on the spatial coherence introduced by the spatial detector. More spatial coherence should be achieved by extending fusion to spatial domain.

The computation time of our method is similar to classical detectors ones. On our test machine (P4, CPU 3 GHz, RAM 2 GB) the average time to process one frame is 2.7 s for our (MC) method, 2.2 s for the SDIp, and 2.7 s for the MRF method. All methods have been penalized by the motion compensation processing time, which is about 2 s for both previous and next frames.

6. Conclusion

In this paper, we proposed a fusion scheme for the detection of blotches in digitized archive film material. This method uses the belief functions framework for combining two blotch detectors, taking advantage of their redundancy, complementarity and incompleteness.

Performance in terms of correct detections and false alarms has been improved, as the decision has been performed after the combination step, taking the conflict between detectors into account. The conflict has been reported on the ignorance mass, thus allowing the introduction of a single risk index. Variable degrees of treatment can be achieved by thresholding this risk index in the decision step.

Further improvements will concern the introduction of discounting factors in order to take into account the detectors reliability, and the introduction of spatial information by spatial fusion. Accurate performance evaluation is still an open issue, because the performance measure should take into account the visual discomfort resulting from the detections missed or from the correction of false alarms.

Finally, this fusion scheme can easily be extended to new detectors, improving detection reliability.

Acknowledgement

This work was supported by the ANRT of the French Ministry of Research and Technology, and by the FP6-IST-507336 PrestoSpace project, supported by the European Commission.

References

Alp, B., Haavisto, P., Jarske, T., Oistamo, K., Nuevo, Y., 1990. Median based algorithms for image sequence processing. SPIE Visual Commun. Image Process., 122–134.

Appriou, A., 1993. Formulation et traitement de l'incertain en analyse multi-senseurs. Quatorzième Colloque GRETSI, Juan les Pins, France, pp. 951–954.

Arce, G., 1991. Multistage order statistic filters for image sequence processing. IEEE Trans. Signal Process. 39, 1146–1163.

Bornard, R., 2002. Probabilistic approaches for the digital restoration of television archives. PhD Thesis, Ecole Centrale, Paris.

Buisson, O., Besserer, B., Boukir, S., Helt, F., 1997. Deterioration detection for digital film restoration. In: IEEE Internat. Conf. on Computer Vision and Pattern Recognition, vol. 1, Puerto Rico, USA, pp. 78–84.

Decenciere Ferrandiere, E., Serra, J., 1997. Detection of local defects in old motion pictures. In: VII National Symposium on Pattern Recognition and Image Analysis, Barcelona, Spain, April, pp. 145–150.

Dromigny-Badin, A., 1998. Fusion d'Images par la Théorie de l'Évidence en vue d'Applications Médicales et Industrielles. PhD Thesis, Institut National des Sciences Appliquées de Lyon, France.

Dubois, D., Prade, H., 1988. Representation and combination of uncertainty with belief functions and possibility measures. Comput. Intell. 4, 244–264.

Gangal, A., Dizdaroglu, B., 2006. Automatic restoration of old motion picture films using spatio-temporal exemplar-based inpainting. In: LNCS 4179, 0055. Springer-Verlag, pp. 55–56.

Gangal, A., Kayikcioglu, T., Dizdaroglu, B., 2004. An improved motion-compensated restoration method for damaged color motion films. Signal Process.: Image Commun. 19, 353–368.

Hamid, M.S., Harvey, N.R., Marshall, S., 2003. Genetic algorithm optimization of multidimensional grayscale soft morphological filters with applications in film archive restoration. IEEE Trans. Circuits Systems Video Technol. 13 (5), 406–416.

Hardie, R., Boncelet, C., 1993. LUM filters: A class of rank-order based filters for smoothing and sharpening. IEEE Trans. Signal Process. 41, 1061–1076.

Kokaram, A.C., 1993. Motion picture restoration. PhD Thesis, Cambridge University, England, May.

Kokaram, A.C., 1998. Motion Picture Restoration: Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video. Springer-Verlag, London, England.

Kokaram, A.C., Rayner, P.J.W., 1993. A system for the removal of impulsive noise in image sequences. In: SPIE Visual Communications and Image Processing, Boston, USA, pp. 322–331.

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nedler-Mead simplex method in low dimensions. SIAM J. Optim. 9 (1), 127–147.

Lefevre, E., Colot, O., Vannoorenberghe, P., 2002. Belief function combination and conflict management. Inform. Fusion, 149–162.

Morris, R.D., 1995. Image sequence restoration using Gibbs distributions. PhD Thesis., Cambridge University, United Kingdom.

Nadenau, M.J., Mitra, S.K., 1996. Blotch and scratch detection in image sequences based on rank ordered differences. In: Fifth International Workshop on Time-varying Image Processing and Moving Object Recognition, Florence, Italy, pp. 1–7.

Nieminen, A., Heinonen, P., Neuvo, Y., 1987. A new class of detail-preserving filters for image processing. IEEE Trans. Pattern Anal. Machine Intell. 9 (1), 74–90.

Rares, A., Reinders, J.T., Biemond, J., 2001. Statistical analysis of pathological motion areas. In: IEE Seminar on Digital Restoration of Film and Video Archives, London, United Kingdom.

Ren, J.C., Vlachos, T., 2007. Efficient detection of temporally impulsive dirt impairments in archived films. Signal Process. 87 (3), 541–551.

Van Roosmalen, P.M.B., 1999. Restoration of archived film and video. PhD Thesis, Delft University of Technology, The Netherlands.

Schallauer, P., Pinz, A., Haas, W., 1999. Automatic restoration algorithms for 35 mm film. J. Comput. Vision Res. 1 (3), 59–85.

Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press.

Smets, P., 1990. The combination of evidence in the transferable belief model. IEEE Trans. Pattern Anal. Machine Intell. 12, 447–458.

Storey, R., 1985. Electronic detection and concealment of film dirt. SMPTE J., 642–647.

Thomas, G.A., 1987. Television motion measurement for datv and other applications. Technical Report, The British Broadcasting Corporation Research Department Engineering Division.

Tilie, S., Laborelli, L., Bloch, I., 2006. Blotch detection for digital archives restoration based on the fusion of spatial and temporal detectors. Fusion 06, Florence, Italy.

Vincent, L., 1993. Greyscale area openings and closings, their efficient implementation and applications. In: Mathematical Morphology and its Applications to Signal Processing, Barcelona, Spain, pp. 22–27.

Wilkinson, M.H.F., Roerdink, J.B., 2000. Fast morphological attribute operations using Tarjan's union-find algorithm. In: International Symposium on Mathematical Morphology, Palo Alto, USA, pp. 311–320.

Yager, R.R., 1987. On the Dempster–Shafer framework and new combination rules. Inform. Sci. 41, 93–138.