Local Picture-repetition Mode Detector for Video De-interlacing

Piedad Brox, Leon Woestenberg, and Gerard de Haan, *Senior Member*, IEEE

Abstract— *The de-interlacing of video material converted from film can be perfect, provided it is possible to recognize the field-pairs that originate from the same film image. Various so-called film-detectors have been proposed for this purpose, mainly in the patent-literature. Typically, these detectors fail in cases where video overlays are merged with film material, or when non-standard repetition patterns are used. Both problems occur frequently in television broadcast. For these hybrid and/or irregular cases, we propose a detector that can detect different picture-repetition patterns locally in the image. This detector combines fuzzy logic rules and spatio-temporal prediction to arrive at a highly robust decision signal, suitable for pixel- accurate de-interlacing of hybrid and irregular video material. In addition to an evaluation of the performance, the paper also provides a complexity analysis.*

Index Terms— Picture-repetition Mode Detection, Video De- interlacing, Pull Down, Video Signal Processing, Fuzzy Inference Systems.

I. INTRODUCTION

Knowledge of the picture repetition pattern is highly relevant for several video signal processing tasks, like video compression, picture-rate conversion and de-interlacing. As this information is usually not included in the transmission, the detection of picture repetition from the video data is necessary. We shall focus on the de-interlacing application [1], which is particularly relevant, since if the field-pairs that are originated from the same image are recognized then the de-interlacing of video material can be *perfect*.

As a large percentage, often the majority, of broadcast video material has been converted from film, methods to realize film- mode detection are currently in demand. In this conversion picture repetition is required, since video signals originating from a video camera provide a picture rate of 50 Hz, or 60 Hz, whereas if the material was registered with a cine-camera the picture rate is only 24 images per second. In order to adapt film to both standard transmissions, a process called *'pull-down'* is performed. Basically, it consists of repeatedly scanning a film image until it is time to show the next. For 25 Hz film shown in a 50 Hz

P. Brox is with the Instituto de Microelectrónica de Sevilla (CNM-CSIC) and the University of Seville, Spain (e-mail:brox@imse.cnm.es). L.M.P. Woestenberg is with the Axon Digital Desing, Udenhout, the Netherlands (e-mail:leon.woestenberg@axon.tv). G. de Haan is with the Philips Research Eindhoven, Eindhoven, the Netherlands. He is also professor at the Information and Communication Group, Technology University of Eindhoven, the Netherlands (e-mail:g.de.haan@philips.com).

broadcast, every image is shown twice and the conversion is referred to as 2:2 pull-down. For 24 Hz film shown on a 60 Hz television, film images are shown alternatingly 2 and 3 times, which is the so-called 3:2 pull-down process. Independent of the type of camera and repetition pattern, interlaced video signals transmit only the odd lines of odd images and the even lines of even images.

Different detectors have been proposed to identify the field-pairs originating from the same film image to enable perfect de-interlacing, or proper picture-rate conversion. Among them zero-vector matching detectors have widely been employed by the majority of current film detectors [2]. They try to match the zero motion vectors on a previous field. To perform it, they normally use two kinds of signals: a first to detect the *frame similarity* and a second one to measure the *field similarity*. Based on the analysis of both similarity metrics, control signals are generated. They indicate the mode of the video signal, i.e. video or film, and the type and phase of the film mode, to determine the image's position in the 3:2 or 2:2 pull-down pattern.

Other approaches try to identify jagged edges in frames. This undesirable phenomenon appears when two fields with moving objects, sampled at different moments at time, are merged into a single image. Several proposals of this kind of detectors have been presented in the literature [3], [4].

Another detector based on edge-detection was proposed in [5]. It analyzes the position of edges in the image since if there is a picture repetition of the fields, edges should be at the same spatial position.

Finally, a motion vector based approach has been proposed in [6]. The sum of the length of the motion vectors is evaluated to decide if two fields are identical or not.

Recent advances in the area of film-detection can be divided into two categories. The first ones report on the increased *robustness* of the algorithms, whereas the second ones focus on the detection of the *local* video mode in hybrid video sequences.

An improved robustness is especially relevant, as an in-correct mode decision produces highly annoying artefacts in the de-interlaced video signal. The approach described in [7] reduces the number of wrong decisions due to vertical details using a new difference metric, whereas the proposed method in [8] uses a layered structure to achieve a robustness improvement.

Local detection has been motivated by the increase of TV material that combines images from different origins in a

single field. None of the techniques previously cited can locally detect different modes in a single field, as their output is a single flag for the entire field. They usually compare the *sum* of absolute values of frame and field differences *over the entire field* with a threshold value [2], [7]. This strategy is far from optimal, since the best threshold value strongly depends on the amount of motion and the level of noise in the picture. Moreover, it leaves no options to distinguish the different modes in a single field that occur in hybrid material. To solve this problem, a method for detecting the film mode of individually moving objects within fields is described in [9]. The identification of these objects is performed using segmentation.

Our proposal combines fuzzy logic and spatio-temporal prediction to increase the robustness of the final decision, and also to take a decision locally on a pixel-by-pixel basis. Due to the capacity of the fuzzy logic-based models to perform a non-linear mapping between the input and output space, they are well-known as good interpolators [10]. One example is the method developed in [11], which uses an adaptive de-interlacing process by weighting between 'field insertion' and 'a spatial interpolation algorithm'. The weighting factors are obtained analyzing, as inputs of the fuzzy system, the intra and inter-field signal differences of the current pixel along a set of pre-determined directions.

In this paper, we propose a set of fuzzy IF-THEN rules to take a decision instead of realizing a weighted interpolation. In this novel approach, each rule models heuristic knowledge to identify one of the possible picture-repetition modes on a pixel-by-pixel basis. To make this pixel-accurate detector robust, a set of proposals are presented. Among them, the main novelty is the inclusion of a spatio-temporal prediction scheme inspired on recursive motion estimation [12]. Here, prediction implies that the final decision not only corresponds to the current pixel, but also the decisions in a spatio-temporal neighborhood of the current pixel are considered.

Since the decision is made on a pixel-by-pixel basis, our method can deal successfully with hybrid video material. Moreover, our proposal is not limited to the recognition of the standard repetitions patterns, like the popular 2:2 or 3:2 pull-down patterns for film. This extends its applicability to any irregular, picture-repetition sequence.
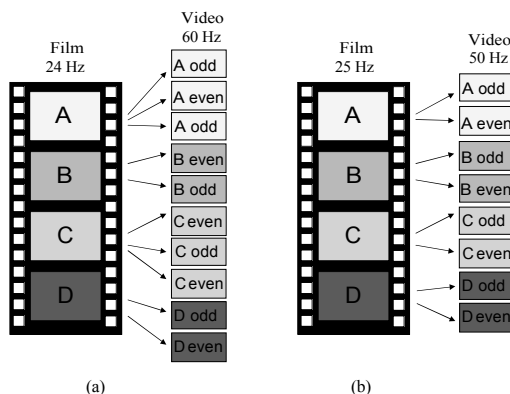
This paper is organized as follows. The proposed algorithm is described in Section II. We present several proposals from a basic one to a more sophisticated one in the different subsections of Section II. The performance of the approach is proven by extensive simulations of video sequences applying the mode detection to perform different de-interlacing techniques. These results are presented in Section III. This section also includes a complexity analysis of the algorithm. Finally, we draw our conclusions in Section IV.

II. LOCAL PICTURE-REPETITION MODE DETECTOR

The proposed mode detector is a decision-making system based on a set of rules. Each single rule models heuristic knowledge to identify locally different modes. As mentioned in the introduction, our proposal, for de-interlacing, offers more than just the functionality of a film-detector, as its rules deal with all possible picture-repetition patterns. To help appreciate the background of the rules, we shall first briefly describe the conversion between film and video, which is still the most common cause of picture-repetition in broadcast video.

The 3:2 pull-down process is common to transfer 24 Hz film to 60 Hz video. To achieve this, every odd film image is scanned twice, while every even film image is scanned three times as shown in Fig. 1(a). Thereafter, the signal is interlaced.

The 2:2 pull-down process is common to transfer 24 Hz film to 50 Hz video. Initially, the picture-rate of the film is increased to 25 images per second by running the film 4% faster. Then, each film image is scanned twice and interlaced, generating two video fields as shown in Fig. 1(b).

To arrive at picture-repetition detection, we calculate three difference signals, the frame difference signal (δ_{frame}) between the next and previous field at the same spatial position (x, y), and the two field differences of the current pixel: with the previous field (δ_{field1}) and with the next field (δ_{field2}). In order to increase robustness against noise the median value of each difference at three vertical positions is used (see Fig. 2). They are defined by the following expressions:

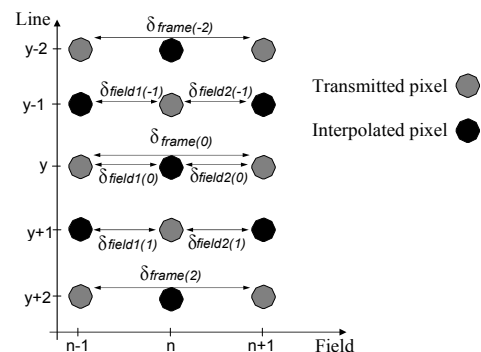$$\delta_{frame}(x, y, n) = med\ (\delta_{frame(-2)}, \delta_{frame(0)}, \delta_{frame(2)}) \qquad (1)$$

Fig. 1. Standard conversion between video and film formats: (a) 3:2 Pull-down. (b) 2:2 Pull-down.

Fig. 2. Picture-repetition mode detector aperture. The shown pixels are used to calculate the local differences.

(a) 3:2 Pull-down

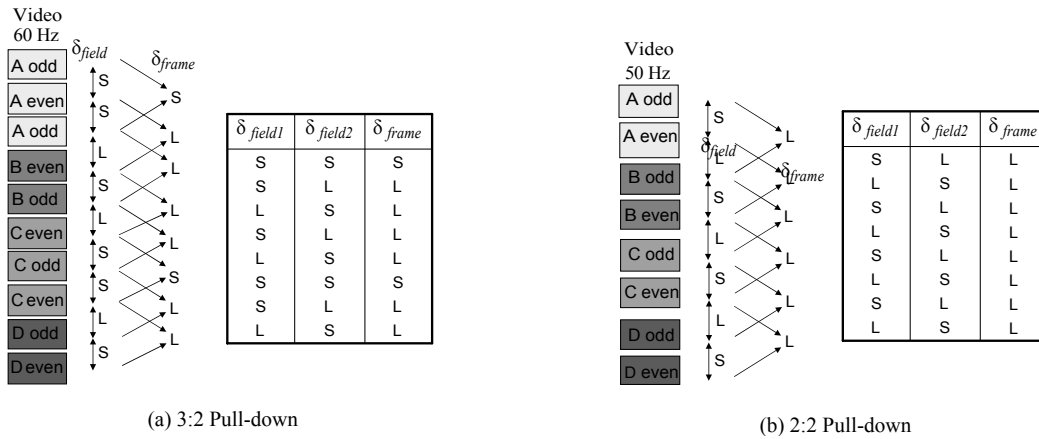| δ_{field1}	δ_{field2}	δ_{frame}
S	S	S
S	L	L
L	S	L
S	L	L
L	S	L
S	S	S
S	L	L
L	S	L

(b) 2:2 Pull-down

δ_{field1}	δ_{field2}	δ_{frame}
S	L	L
L	S	L
S	L	L
L	S	L
S	L	L
L	S	L
S	L	L
L	S	L

Fig. 3.   Temporal difference patterns of standard conversions.

$$\delta_{frame}\ (x, y, n) = med\ (\delta_{frame(-2)},\ \delta_{frame(0)},\ \delta_{frame(2)})\qquad(1)$$

$$\delta_{field1}\ (x, y, n) = med\ (\delta_{field1(-1)},\ \delta_{field1(0)},\ \delta_{field1(1)})\qquad(2)$$

$$\delta_{field2}\ (x, y, n) = med\ (\delta_{field2(-1)},\ \delta_{field2(0)},\ \delta_{field2(1)})\qquad(3)$$

where n denotes the field number in the sequence order:

$$\delta_{frame(i)}\ (x, y, n) = |F\ (x, y + i, n + 1) - F\ (x, y + i, n - 1)|$$

$$\delta_{field1(i)}\ (x, y, n) = |F_d\ (x, y + i, n) - F\ (x, y + i, n - 1)|$$

$$\delta_{field2(i)}\ (x, y, n) = |F_d\ (x, y + i, n) - F\ (x, y + i, n + 1)|$$

To calculate the differences, a simple initial de-interlacing algorithm is used to generate progressive frames (F_d). Typically, a vertical-temporal median or a vertical-temporal linear filter is proposed [1]. If the initial de-interlacing process would be perfect, the difference between fields from the same source image, as it occurs with film, should be equal to zero. With a simple and realistic initial de-interlacing algorithm alias and vertical details in the field may introduce false detections of motion. In order to reduce this problem, the field differences shown in expressions (2) and (3) are normalized by vertical intra-fields differences.

The different types of temporal differences patterns are shown in Fig. 3(a) and 3(b) for the pull-down 3:2 and 2:2 process, where 'L' means a large difference and 'S' a small difference. Considering these temporal difference patterns, the following knowledge can be applied to detect the different modes:

1) If the frame difference is large and both field differences are large too, then the pixel corresponds to a moving object in a video sequence, where all fields are different.
2) If the frame difference is small and both field differences are also small, then the pixel corresponds to an area without motion during these two field-periods. There- fore, it must be, either a stationary area, or a moving area if the 3 fields originate from a 3 times repeated image, e.g. as it occurs with 3:2 pull-down.
3) If the frame difference is large, but one of the field differences is small while the other is large, then the current pixel belongs to a sequence with picture-repetition, where at least one field is repeated, as it occurs e.g. in 2:2 pull-down mode.
4) Otherwise, none of the modes is identified. This the may occur when the initial de-interlacing process, then necessary to calculate the field differences, suffers, from alias, because the signals are corrupted by noise, or because the image has a flat area.

This heuristic knowledge can be modelled using a system with fuzzy IF-THEN rules, since the concepts large and small are understood as fuzzy definitions instead of threshold values. Using fuzzy logic, the concepts of 'SMALL' and 'LARGE' are represented by fuzzy sets, the membership values of which change continuously between 0 and 1, as shown Fig. 4(a) and Fig. 4(b).

Each fuzzy IF-THEN rule in our system has antecedent[1] linguistic values and a single consequent mode as shown in Table I. The minimum/maximum[2] operators are selected as connectives 'and'/'or' of the antecedents, respectively.
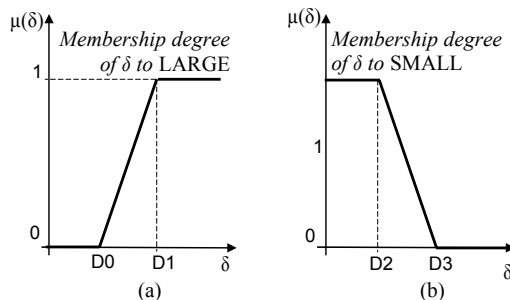

Fig. 4.   Membership functions for the fuzzy sets (a) LARGE, (b) SMALL.

[1] Antecedent is the common term for a condition in the fuzzy logic domain [13]

[2] Minimum and maximum operators are usually defined as 'and' and 'or' operators in the fuzzy logic domain respectively [13]

TABLE I
FU Z Z Y RU L E SE T

if	antecedent	then	consequent
1)	$\delta_{frame}(x,y,n)$ is LARGE and $\delta_{field1}(x,y,n)$ is LARGE and $\delta_{field2}(x,y,n)$ is LARGE		MODE is video
2)	$\delta_{frame}(x,y,n)$ is SMALL and $\delta_{field1}(x,y,n)$ is SMALL and $\delta_{field2}(x,y,n)$ is SMALL		MODE is stationary
3)	($\delta_{frame}(x,y,n)$ is LARGE and $\delta_{field1}(x,y,n)$ is SMALL and $\delta_{field2}(x,y,n)$ is LARGE) or		MODE is repetition
	($\delta_{frame}(x,y,n)$ is LARGE and $\delta_{field1}(x,y,n)$ is LARGE and $\delta_{field2}(x,y,n)$ is SMALL)		
4)	otherwise		MODE is undetermined

The use of the operator 'and' forces the system to analyze the field differences signals only if the corresponding frame difference signal is large. This strategy increases the robustness of the detection, since the frame difference signal is more reliable than the other differences that are based on imperfect initial de-interlacing results (F_d). The main advantage of the fuzzy-logic based approach is that it provides a smooth transition between one decision and another. The activation degree of a rule (α_i) indicates the compatibility grade of the (i^{th}) IF-THEN rule, which is calculated by computing the membership values of the antecedents:

$$\alpha_1(x, y, n) = min(\mu_{LARGE}(\delta_{frame}), \mu_{LARGE}(\delta_{field1}), \mu_{LARGE}(\delta_{field2})) \quad (4)$$

$$\alpha_2(x, y, n) = min(\mu_{SMALL}(\delta_{frame}), \mu_{SMALL}(\delta_{field1}), \mu_{SMALL}(\delta_{field2})) \quad (5)$$

$$\alpha_3(x, y, n) = max((\alpha3a, \alpha3b)) \quad (6)$$

$$\alpha_4(x, y, n) = 1 - \alpha_1 - \alpha_2 - \alpha_3 \quad (7)$$

where:

$$\alpha_{3a}(x, y, n) = min(\mu_{LARGE}(\delta_{frame}), \mu_{LARGE}(\delta_{field1}), \mu_{SMALL}(\delta_{field2})) \quad (8)$$

$$\alpha_{3b}(x, y, n) = min(\mu_{LARGE}(\delta_{frame}), \mu_{SMALL}(\delta_{field1}), \mu_{LARGE}(\delta_{field2})) \quad (9)$$

For each pixel, the values $\alpha_{\{1,2,3,4\}}$ are the output signals of the fuzzy system. Each signal corresponds to the activation degree of an individual rule and ranges from 0 to 1. Since our proposed detector aims at a pixel-by-pixel mode decision, alternative robustness measures are necessary. These are described in the following subsections.

A. Increase of the robustness of the fuzzy system decision

The proposed reasoning method is based on a single winner rule. The winner is the fuzzy IF-THEN rule that has the maximum activation degree, that is, the maximum compatibility grade with one of the patterns
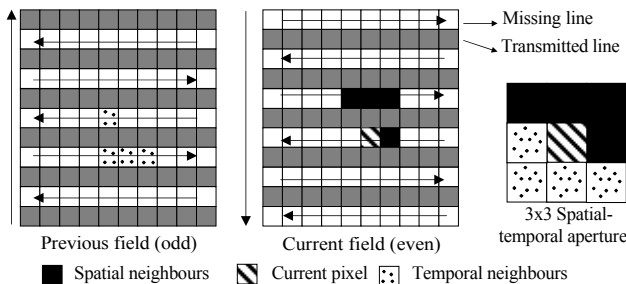

Missing line
Transmitted line
3x3 Spatial-temporal aperture

Previous field (odd)   Current field (even)

■ Spatial neighbours   ◩ Current pixel   ⸭ Temporal neighbours

Fig. 5.   The decisions at 9 positions in a 3x3-aperture are involved in the decision-making process.

described by the antecedents. However if multiple activation degrees of contrary rules are activated, choosing the maximally activated mode easily results in wrong decisions. To improve this, a decision is adopted when its corresponding rule is the most activated and also the activation degree of the contrary rule is very low.

B. Spatio-temporal Prediction

In our proposal, the final decision for the current pixel is not only based on the decision of the system for the current pixel, but applies 'spatio-temporal' predictions, taking into account also the decisions of the pixels in a 3x3 neighborhood.

The idea is to make a decision only when one mode is actually clear, and propagate the decision until a new clear decision is taken. To reduce error propagation a meandered scanning is proposed (see Fig. 5). The detector processes the even fields in a streaming fashion, that is, from the top-left pixel to the bottom-right pixel, whereas the odd fields are processed from the bottom-right pixel to the top- left pixel . Four pixels in the 3x3 window are spatial neighbors and belong to the current field, and four are temporal neighbors of the previous field as shown Fig. 5.

C. Temporal forward prediction process

The temporal predictions are more complex than the spatial prediction. To illustrate the problem, let us consider the temporal difference pattern of pull-down 3:2 process shown in Fig. 3(a). Analysis of this pattern shows that not only the third rule is activated but also the second one. To be exact, the second rule is activated every five fields of the video sequence. This means that the MODE of a pixel with the same spatial coordinates in the previous field not always has to agree with the current one. For any picture repetition pattern, decisions from the previous field can be transformed into new predictions as shown in Fig. 6. The values of difference signals are represented using the notation 'LSL', which means a LARGE difference of δ_{field1} and a SMALL difference of δ_{field2} and a LARGE value of δ_{frame} . For each value of the difference signals in the previous field, the different alternatives for a pixel in the current field are shown in Fig. 6. From the knowledge of the previous mode, only the value of δ_{field1} can be assigned. For instance, a pixel from the previous field where video mode is detected implies a large value of difference signals, that is 'LLL'. In this case, the value of δ_{field1} will be surely 'L' in the next field and then, δ_{frame} will be also 'LARGE'. However, there is no information to predict the value of δ_{field2}. Analyzing each possibility the

Previous difference values	Updated difference values
LLL	L?L
SSS	S??
LSL	S??
SLL	L?L
???	???

Mode of a pixel in the previous field	Temporal prediction of the Mode
MODE is video	MODE can be video or repetition
MODE is stationary	MODE can be stationary or repetition
MODE is repetition	MODE can be video, repetition and stationary
MODE is undetermined	MODE is undetermined

Fig. 6. Temporal predictions for the MODE in the previous field.

Fig. 8. If picture-repetition is assumed incorrectly, feathering (left) will show.

temporal predictions for the modes are shown in Fig. 6. The multiple modes of the temporal predictions have an equal bias.

D. Reinforcement of the final decision

The flowchart of the process is shown in Fig. 7. The input signal is the MODE of the pixels in the 3x3 neighborhood. Only if the occurrence of modes exceeds a set of values $C_{\{1,2,3,4\}}$ and there are no undetermined decisions, a control signal called PATTERN is activated. If not, i.e. the majority of decisions are undetermined, the control signal will be generated to code the PATTERN signal for 'undetermined' decisions. Finally, in pixels where none of the rules is sufficiently activated, the decision of the previous pixel in the scanning directions is assigned.

Some erroneous mode decisions are more critical than others. For our de-interlacing application, repetition mode leads to perfect results through merging the lines of fields that belong to the same film image. However, when this mode is erroneously detected, e.g. in video camera material, annoying feathering artefacts appear in the de-interlaced picture, as shown in Fig. 8. The consequences of the video mode decision is less critical, since in this case a robust de-interlacing technique is employed that performs sub-optimal on film material, but gives not very objectionable artefacts. Given this asymmetrical behaviour, priority is given to the video-mode, i.e. the mode for which we assume there is
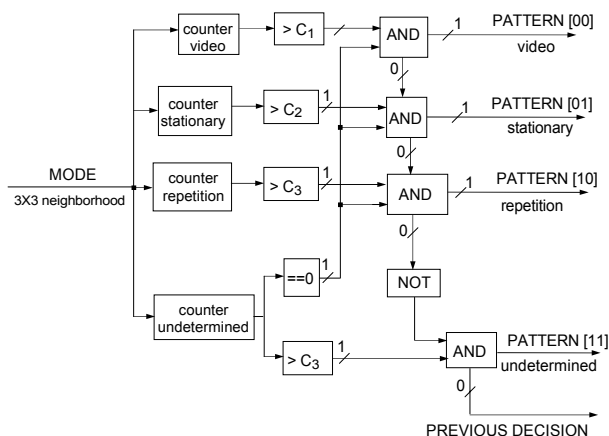
no picture repetition. Therefore, this mode is corroborated firstly and it requires a lower number of modes in the aperture to take a decision: C1 <C2≈C3≈C4 (see Fig. 7).

Through analysis of the two parts of the third rule, a second control signal is generated named 'PHASE' to identify which two of the three fields are identical in the case that a picture repetition, e.g. a repetition mode, is detected. Fig. 9 shows the placement of the fuzzy mode detector in the video processing chain for our de-interlacing application. Both signals generated by the fuzzy system are used as control signals to determine the de-interlacing strategy. In case that a repetition mode is detected, the de-interlacing process becomes perfect by weaving two fields together. This is also valid in the case that the pixel is classified as belonging to a stationary area. However, if the first or the fourth rule are the most activated, a conventional video de-interlacing approach [1] has to be used.

E. Improving the performance of the fuzzy system by membership function learning

From heuristic knowledge, there is no restriction to fix the parameters of membership functions. However, some values will provide better results than other ones. Our idea is to select the most suitable values using a set of input/output training patterns of image sequences where the mode decision is known by the designer. Five fields of two different sequences were used. Originally both sequences are video material but the 3:2 and 2:2 pull-down cadences were generated and used as training patterns. These sequences are called Kielp and Bicycle in Section III. Simulation results in Section III prove
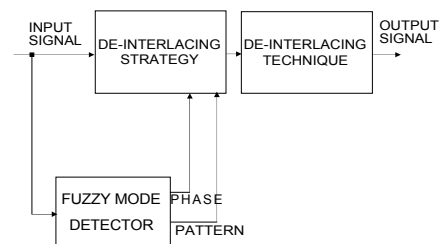

Fig. 7. Flowchart of the decision block to increase the robustness of the control 'MODE' signal.

Fig. 9. Placement of the fuzzy film-mode detector in the video chain.

TABLE II
MEMBERSHIP FUNCTION PARAMETERS AFTER THE LERANING PROCESS

Input variable	Parameters	Mode video	Mode ' stationary'	Mode 'repetition'		Mode 'undetermined'	Initial parameters
				LSL	SLL		
δframe	D0	1	0	0.5	0	0	0
	D1	8	8	1.5	8	8	8
	D2	0	0	0	0	0	0
	D3	8	8	8	8	8	8
δfield2	D0	0	0	0	0	0	0
	D1	2	2	2	2	2	2
	D2	0	0	0	0	0	0
	D3	2	2	2	2	2	2
δfield1	D0	0	0	0	0	0	0
	D1	0.25	2	0.5	2	2	2
	D2	0	1	0	1	0	0
	D3	2	8.5	2	9	2	2

that the method is robust for a wide number of test sequences.

For the tuning process, we used the development environment Xfuzzy3.0 [14]. This is an environment for designing fuzzy sets that is composed of a set of CAD tools covering the different stages of description, verification, simplification and synthesis of inference systems based on fuzzy logic. Xfuzzy3.0 integrates a CAD tool, named *xfsl* [15], to tune fuzzy systems described in the environment.

We further applied a set of training video sequences. Only the values D0, D1, D2 and D3 that define the membership functions have been adjusted in the learning stage. The Levenberg-Marquardt algorithm has been selected as supervised learning algorithm and the results of the process are shown in Table II. This table shows the tuned parameters for each one of the modes, and also the initial parameters of the membership functions that were fixed manually. For the first field of the video sequence, the initial parameters are used. For the rest of the fields, the tuned parameters for each mode is taken.

F. Mode filtering to improve robustness

Since an erroneous video detection in the repetition area is less serious than an erroneous repetition mode decision, a simple spatial filtering is performed to spread the video mode decision. Fig. 10 shows the shape of the spatial aperture. As it can be seen, it contains more pixels in the vertical direction than in the horizontal. The reason is that the image is
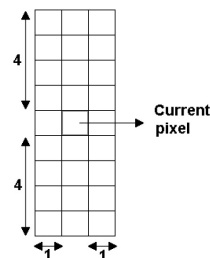
processed in a streaming direction, so mistakes are transmitted along horizontal direction. To avoid this, a higher number of pixels in vertical direction are considered. The final structural overview of the proposed detector is shown in Fig. 11.

III. PERFORMANCE OF THE PROPOSED ALGORITHM

The performance of the proposed algorithm has been evaluated in the de-interlacing application. We investigated the image quality and calculated the computational cost of the detector. Subsection A describes the cost calculations. A brief description of the video test sequences can be found in Subsection B and finally, the overall performance is given in Subsection C.

A. Algorithm Cost

The algorithmic cost is measured using the number of floating point operations (FLOPS) as a(n inverse) figure of merit. The algorithm requires 543.7 Megas floating point operations to analyze one field of a video sequence with a resolution of 720x576. We have considered this
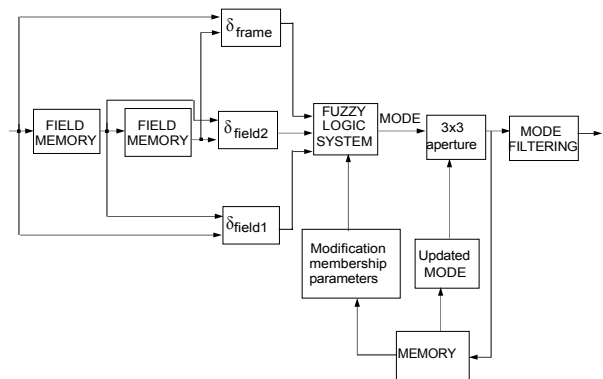

Fig. 10.   3x9-aperture for the spatial filtering.

Fig. 11. Structural overview of the film detector.

(a) TMF sequence   (b) Fire-rose sequence   (c) Renata hybrid sequence
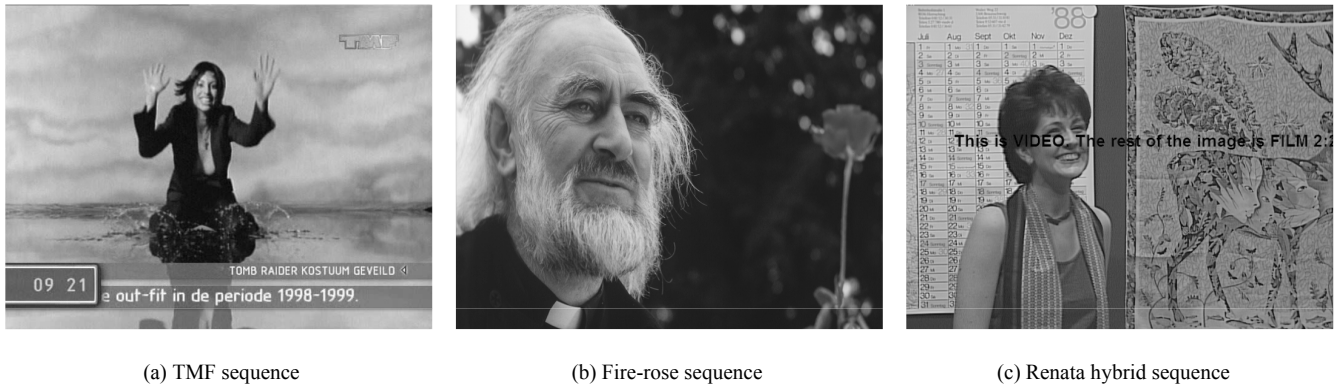
Fig. 12. Snapshots of real sequences used to prove the performance of the proposed algorithm.

measurement instead of computational time as it strongly depends on the platform on which the algorithm is implemented and the efficiency of the programming.

B. Description of the Sequences

Subsection C contains results from the analysis of several sequences. Some of them are real sequences from TV channels or movies, and others are test video material. Forty fields of each sequence have been processed. Three of these sequences have been especially analyzed:

- TMF. This is an original sequence captured from a Dutch broadcast channel called TMF. The sequence is an interlaced video clip (2:2 pull-down mode) with an overlay containing a ticker-tape video text as can be seen in Fig. 12(a). It also contains stationary areas (around the clock and the TMF-logo).

- Fire-rose. This is an interlaced 2:2 film sequence. The detection of repetition mode is difficult due to the fine details in the man's beard as shown in Fig. 12(b). Moreover this sequence contains a very low level of motion.

- Renata. This sequence has been used to show the improvements introduced by the individual robustness measures explained in Section II. It is originally a video scene. However, it has been artificially transformed into an interlaced 2:2 repetition mode. The sequence has then been converting into a hybrid sequence, by adding a horizontally horizontally moving video text in the middle of the fields, as shown in Fig. 12(c).

C. Simulation Results

To prove the performance of the proposed detector, the three-fields VT filtering approach [16] is used if the 'video' or 'undetermined' mode is detected. On the other hand, if one of the 'repetition' mode is detected, the de-interlacing process is implemented by weaving. Comparing the de-interlaced with the original progressive picture of Renata, a Mean Squared Error can be calculated. Fig. 13 shows the relative MSE-score as a percentage of the MSE-score obtained with the VT filtering [16]. As can be seen, our final proposal

reduces the total MSE error with almost 60 %. It includes the improvements that are described in subsections {A, B, C, D, E, F} of Section II. The results achieved by the proposal with a modification of membership function parameters are slightly better (column P4 in Fig. 13) than the obtained with fixed parameters (column P3 in Fig. 13).

The detector has also been used to de-interlace the real sequences shown in Tables III and IV. It decreases the total MSE score by a high factor in percentage (almost 100%) in the majority of film sequences (see Table III). This not only produces a perfectly de-interlaced image, but also considerably reduces the complexity as weaving is the method with the lowest computational cost. Due to the presence of low motion and/or a high number of the details, repetition mode is not well detected in some of the film sequences and the MSE only falls to 40%. This is not crucial for de-interlacing applications because conventional de-interlacing is applied when actual repetition is misinterpreted as video.

Finally, the total MSE is slightly reduced when the detector is used for video sequences (see Table IV). This is due to the improved de-interlacing of the few static areas. Although the modification of membership function parameters does not introduce many advantages for the Renata hybrid
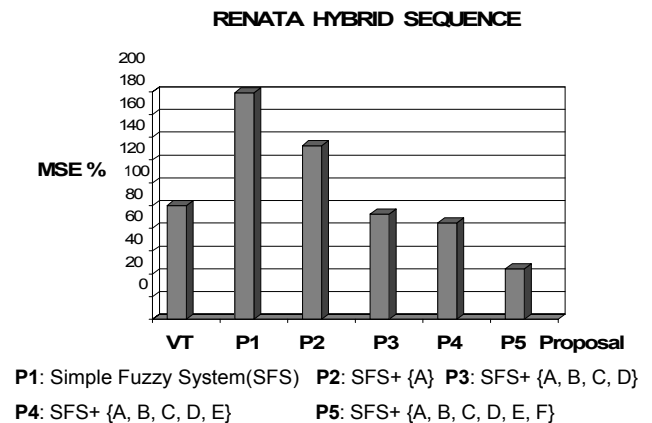

P1: Simple Fuzzy System(SFS)   **P2**: SFS+ {A}   **P3**: SFS+ {A, B, C, D}
P4: SFS+ {A, B, C, D, E}   **P5**: SFS+ {A, B, C, D, E, F}

Fig. 13. MSE percentage of each one of the proposals.

TABLE III
SIMULATION RESULTS OF FILM SEQUENCES

Sequence	Shrek	Gladiator	Stairs	Fargo office	Fargo repair	Matze1	Matze2	Fire rose	Chop hunt	Fargo	Vanessa	Chop land	Flight
Total MSE													
Detector off	327.607	62.72	93.96	208.54	327.31	745.74	682.17	204.61	52.01	364.13	217.53	220.69	231.65
Detector on	1.52	0.098	0	124.77	0.143	419.78	0	0	0.077	0.74	0.21	2.83	0
Reduction(%)	99.53	99.84	100	40.16	99.95	43.708	100	100	99.85	99.79	99.901	98.71	100

TABLE IV
SIMULATION RESULTS OF VIDEO

Sequence	Animatrix-a	Animatrix-b	Dieanotherday	Bicycle	Kielp	Girlgate	Wman	Renata	Xmen2	Newsreader
Total MSE										
Detector off	139.71	712.46	446.38	1517.03	5321.83	156.109	71.56	450.425	266.52	1027.992
Detector on	52.06	711.91	371.29	1503.52	5319.12	154.71	71.45	449.79	250.81	1027.93
Reduction(%)	62.74	0.03	16.83	0.9	0.05	0.8	0.19	0.01	5.89	0.01

sequence, it is necessary to achieve good results for some sequences in Tables III and IV.

The response of the detector was also analyzed for the TMF and Fire-rose sequences. The output modes when processing the snapshot in Fig. 12(a) can be visually corroborated in Fig. 14(a). In this figure, white color means repetition mode, light g grey means stationary areas, dark grey corresponds to video mode and black color shows zones where the decision is not clear. As can be seen, the critical areas of the field are correctly detected. The MSE value for the TMF sequence could not be included in Table III since the original progressive material is not available.

The performance of the original Fire-rose (film material) sequence is perfect as shown in Table III. If this sequence is transformed into video by eliminating the repeated fields,

video mode is also correctly detected despite the low level of motion as shown Fig. 14(b).

Finally, a test is proposed to prove the advantages of using fuzzy definitions of the concepts SMALL and LARGE instead of crisp definitions. The results show a more critical distinction among the different mode areas of the field if crisp definitions are used. This produces serious mistakes as it can be seen in Fig. 15(a) for Renata hybrid sequence and in Fig. 15(b) for TMF sequence.

Unfortunately, there is no competitive detector that performs a local picture repetition mode detection in current scientific literature. This is why comparisons with other proposals of similar characteristics are not included in this section.

IV. CONCLUSIONS

The de-interlacing of video material converted from film can be perfect, provided it is detected correctly. Typically, however, available detectors fail in cases where video overlays are merged with film material, or when non-standard repetition patterns are used. Both problems occur frequently in television broadcast. For these hybrid and/or irregular cases, we have proposed a detector that is capable to detect locally in the image different picture-repetition patterns. By distinguishing only the following cases:

- Stationary, i.e. all 3 fields show object at same position
- No identical fields, i.e. all 3 fields show object at different position

Paired identical fields case A, first two fields show object at same position, third field at different position
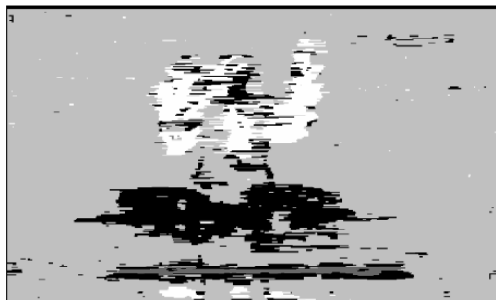
(a)
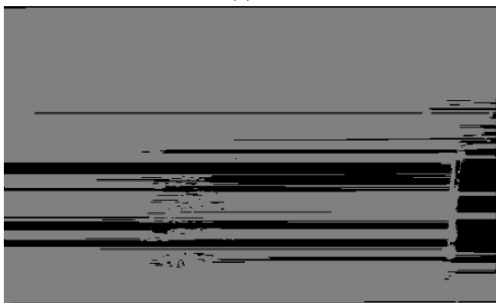
(b)

Fig. 14. Mode decisions taken by the system for (a) TMF and (b) Renata sequences. White indicates repetition, light grey stationary, dark grey video and black are unclear areas.

(a) Renata sequence   (b) TMF sequence

Fig. 15. Simulation results using crisp definitions of LARGE and SMALL.

Paired identical fields case B, last two fields show object at same position, first field at different position

- Unclear, i.e. the local data is ambiguous we designed a picture repetition detector, suitable for all possible patterns without limitation to the common patterns, like 2:2 and 3:2 pull-down.

For instance, a long arbitrary cadence such as 3:2:2:3 can be detected since the rules antecedents only compute absolute differences among three consecutive fields. The detector combines fuzzy logic rules, to deal with uncertain cases, and uses spatio-temporal prediction to get a robust decision signal even in unclear areas. Our evaluation shows a very favourable performance and an attractive low computational complexity.

REFERENCES

[1] G. de Haan and E. B. Bellers, De-interlacing - an overview. Proc. of the IEEE, vol. 86, Issue 9, pp. 1839-1857, Sep. 1998.

[2] T. C. Lyon and J. J. Campbell, Motion sequence pattern detector 1, 1991. United States Patent Office US 4,982,290.

[3] C. Correa and R. Schweer, Film mode detection procedure and device. Assignee: Deutsche Thomson-Brandt GMBH, Villingen-Schwennigen (DE), July 1, 1998. European Patent Office 0567072B1.

[4] H. Y. W. Lucas, Progressive/interlace and redundant field detection for encoder. Applicant: STMICRO-Electronics Asia Pacific PTE LTD, Singapore, June, 2000. World Intellectual Property Organization, International Publication Number: WO 00/33579.

[5] P. Swan, System and method for reconstructing noninterlaced captured content for display on a progressive screen. Assignee: ATE Technologies, Inc. Thornhill, anada, Apr. 25, 2000. United States Patent Office US 6,055,018.

[6] G. de Haan, H. Huijgen, P .Biezen, and O. Ojo, Method and apparatus for discriminating between movie film and non-movie film and generating a picture signal processing mode control signal. Assignee: U.S. Philips Corporation, New York, USA, Nov. 15, 1994. United States Patent Office US 5,365,280.

[7] A. Dommisse, Film detection for advanced scan rate converters, M. Sc.Thesis, TUE, Eindhoven, Aug. 2002.

[8] C.-C. Ku and R.-K. Liang, Robust Layered Film-Mode Source 3:2 Pulldown Detection/Correction. IEEE Trans. on Consumer Electronics, vol.50, no.4, pp. 1190-1193, Nov. 2004.

[9] G. de Haan and R. B. Wittebrood, Recognizing film and video object occuring in parallel in single television signals fields. Assignee: Konin- klijke Philips Electronics N. V., Eindhoven, NL, Aug. 30, 2005. United States Patent Office US 6,937,655.

[10] J. L. Castro, Fuzzy logic controllers are universal approximators. IEEE Trans. on ystems, Man and Cybernetics, vol.25, no.4, pp. 629-635, Apr. 1995.

[11] L. He and H. Zhang, Motion object video on film detection and adaptive de-interlace method based on fuzzy logic. Assignee: nDSP Corporation, Campbell, CA, Sep. 28, 2004. United States Patent Office US 6,799,168.

[12] G. de Haan, and P. W. A. C. Biezen, Sub-pixel motion estimation with 3-D recursive search block-matching.Signal Processing: Image Communication 6, pp.229-239, June 1994.

[13] E. Cox, The Fuzzy Systems Handbook: A Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems. AP Professional Editorial (2nd edition).

[14] F.J. Moreno-Velo, I. Baturone, S. Sánchez-Solano, and A. Barriga, Rapid design of complex fuzzy systems with Xfuzzy.Proc. IEEE Int. Conf. on Fuzzy Systems, pp.342-347, St. Louis, USA, May 2003.

[15] F.J. Moreno-Velo, I. Baturone, R. Senhadji, and S. Sánchez-Solano, Tuning complex fuzzy systems by supervised learning algorithms.Proc. IEEE Int. Conf. on Fuzzy Systems, pp.226-231, St. Louis, USA, May 2003.

[16] M. Weston, Interpolating lines of video signals. Assignee: British Broadcasting Corporation, London, GB2, Dec. 6, 1988. United States Patent Office US 4,789,893.

Piedad Brox was born in 1979 in Córdoba, Spain. She received the degree in Electronic Physics from the University of Córdoba in 2002, and the Advanced Studies Degree in Microelectronics from the University of Seville in 2004.

Since 2002, she has been with the Microelectronics Institute of Seville (IMSE), which belongs to the Spanish Research Council (CSIC) and the University of Seville. She received a fellowship of Introduction to Research for Undergraduate Students during 2002.

Currently, she is a Postgraduate Re- search Fellow and has a Postgraduate Fellowship under F.P.U. program from the Spanish government. Her research areas are fuzzy image processing algorithms and their hardware implementation (starting from high-level descriptions to prototyping in FPGAs). She belongs to the 'Digital and Mixed Signal Integrated Circuit Design Group' of IMSE.

Leon Woestenberg received the M.Sc. degree in Electrical Engineering from the University of Tech- nology Eindhoven in 2006. His final project involved content-adaptive video de-interlacing under supervision of Gerard de Haan.

Since 2000 he has been employed, and performed his M.Sc. final project at Axon Digital Design, a manufacturer of audio and video processing systems for television broadcast and production. As an embedded software engineer he developed and re-used open-source software in a commercial environment.

Since 1992 he is involved in open-source software projects as developer and project manager. Currently, since 2004, as a senior systems designer he designs signal processing and control systems and software and acquires and prepares newly available technology for use in future designs.

His professional interests involve operating systems, hardware/software codesign, video and graphics, the open-source development model and the use of open-source software in commercial environments.

Gerard de Haan (Senior Member, IEEE) received B.Sc., M.Sc., and Ph.D. degrees from Delft University of Technology in 1977, 1979 and 1992 respectively. He joined Philips Research in 1979. He has led research projects in the area of video processing, and participated in European projects. He has coached students from various universities, and teaches since 1988 for the Philips Centre for Technical Training. Currently, he is a Research Fellow in the Video Processing & Visual Perception group of Philips Research Eindhoven, and a part-time full Professor at the Eindhoven University of Technology teaching "Video Processing for Multimedia Systems".

He has a particular interest in algorithms for motion estimation, video rate conversion, and image enhancement. His work in these areas has resulted in several books, more than 120 papers, about 100 patents, and various commercially available ICs. He was the first place winner in the 1995 and 2002 ICCE Outstanding Paper Awards program, the second place winner in 1997, 1998 and 2003, and the 1998 recipient of the Gilles Holst Award. In 2002, he received the Chester Sall Award from the IEEE Consumer Electronics Society.

The Philips 'Natural Motion Television' concept, based on his PhD-study received the European Video Innovation Award of the Year 95 from the European Imaging and Sound Association. In 2001, the successor of this concept "Digital Natural Motion Television" received a "Business Innovation Award" from the Wall Street Journal Europe. Gerad de Haan is a Senior Member of the IEEE.