MASTER

Film detection for advanced scan rate converters

Dommisse, A.

Award date:
2002

TU/e technische universiteit eindhoven

Faculty of Electrical Engineering
Section Design Technology For Electronic Systems (ICS/ES)
ICS-ES 807

Master's Thesis

FILM DETECTION FOR ADVANCED SCAN RATE CONVERTERS.

A. Dommisse

Coach: Ing. A.K. Riemens (Philips Research Laboratories)
Dr. C. Henschel (Philips Research Laboratories)
Supervisor: prof.dr.ir. G. de Haan
Date: August 2002

Film Detection for Advanced Scan Rate Converters

A. Dommisse

August 20, 2002

Abstract

Video signals originating from video-cameras have properties that differ from video originating from film-scanners. For advanced scan-rate conversion, it is important to distinguish between the two types and adapt the processing to the type. Incorrect processing in such scan-rate converters could lead to undesirable artifacts in the converted video signal. Therefore a robust method is required to classify video signals, based on their origin. Since this method applied in consumer electronics, such as high-end televisions, a cheap implementation is also required. In this report such detection mechanisms, better known as film detectors, are discussed.

Based on the available literature, an overview of the field of film detection is given. The Schutten-Riemens Film Detector (SRFD) is examined.

Results include:

- A new difference metrics is proposed. This difference metric is insensitive to static vertical detail and has only a 8.5% increase in cost, compared to the original SRFD.
- A hard cut shot boundary detection scheme is implemented and tested.
- A new film mode classification is proposed: The *hybrid mode*. This mode describes mixed video sequences that have both the characteristics of video mode and film mode.
- A new source of mixed mode sequences has been identified. The encoding process in DV-cameras produces a motion pattern that can be detected as 2:2 pull-down film mode by the film detector.

Contents

Preface

This document is my Master's thesis and as such the result of a graduation project to obtain the Master of Science degree in Electrical Engineering from the Eindhoven University of Technology. The proposal of this project initiated at Philips Research In Eindhoven and the major part of the work has been performed there in the Video Processing and Visual Perception (VIPs) Group.

The report assumes a fair knowledge of video processing and video signals. I recommend a book that gives an excellent overview of this subject field: 'Video Processing for multimedia systems' by G. de Haan [17].

This research described in this report is far from complete. This report is snapshot of the current research. Despite of this I hope that you enjoy reading this report just as much as I have enjoyed writing it.

I wish to thank Philips research for giving me the opportunity to perform this assignment. I would also like to thank Laurens Doornhein and Jeroen Kettenis for supplying information about the Bendic and Prozonic video processors and Frits de Bruijn, Gerard de Haan, Christian Hentschel, Radu Jasinschi, Matthijs Piek, Bram Riemens, Robert-Jan Schutten, Rimmert Wittebrood and the rest of the VIPs group at Philips Research Eindhoven for their support and input.

Chapter 1

Introduction

This report deals with the field of video processing. More specifically the area of film detection or video source detection for application in advanced video format converters. Video-format converters are devices that convert video signals from on one video format to another video format. A typical example is the de-interlacer in high-end televisions. A de-interlacer converts an interlaced video signal into a non-interlaced, i.e. progressively scanned, video signal.

Film detection is required for video-format converters for:

- The removal of the motion judder artifact [17]. This artifact is introduced when motion picture film is converted into a video signal with conventional equipment.
- Perfect de-interlacing using the field insertion method.

In both cases it is important to know whether the source of the video signal is a motion picture film converted to video, or a video camera. The means to obtain this knowledge is called a film detector.

In the video chain (Figure 1.1), the film detector analyses the incoming video signal. From this analysis, control signals are generated. The video-format converter adapts the processing of the incoming video signal based on these control signals. This adaptation increases the perceived image quality of the video signal.

However, an erroneous detection of the source can lead to severe artefacts. These artefacts (figure 1.2) can occur when a video signal originating from a video camera is processed as a converted motion picture. For this reason, the accuracy of the film detector is one of its most important properties.

The control signals indicate the 'mode' of the video signal. This 'mode' is used for registering the image data. In this stage we define two modes:

Video mode Image data that is registered with a video camera. Each field in the video signal contains image data from a unique temporal instance.

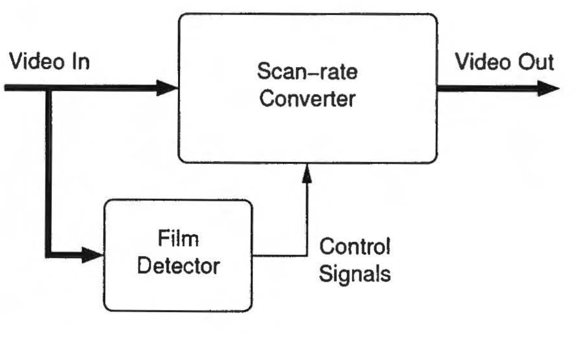


Figure 1.1: A (part of a) video chain, consisting of a film detector and a scan-rate converter

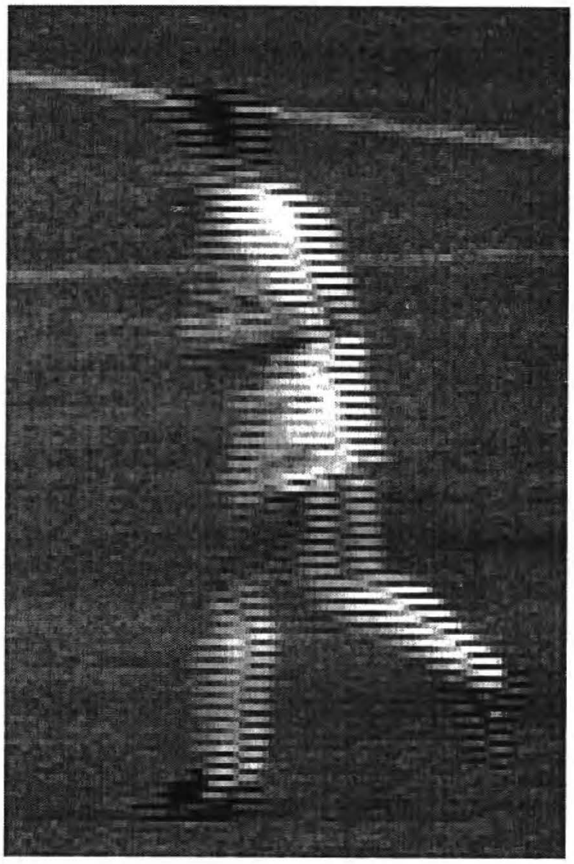


Figure 1.2: An incorrectly de-interlaced frame. The sequence was in video mode, but was treated as a sequence in film mode, introducing severe artifacts.

Film mode Image data that is registered with a film camera and is converted into a video signal. Images are repeated using a sample and hold scheme, to compensate for the lower picture rate. Different fields contain image data from the same temporal instance.

In film mode, the control signals also indicate whether the current image is a repeat of a previous image. Using this information, the scan-rate converter adapts its processing to increase the quality of the out-going video signal.

1.1 Video and film standards

Until 1990 video standards were designed, according to de Haan [17] to strike a particular compromise between quality, cost, transmission or storage capacity and compatibility with other standards.

According to de Haan [17, page 103], when focusing on picture rates, three formats can be distinguished:

50 Hz video A transmission standard, commonly known as PAL or SECAM[1] [21], that consists of 50 interlaced fields per second. Each frame consists of 625 lines of which the even and odd lines are alternated transmitted as fields.

60 Hz video A transmission standard, commonly known as NTSC[2], that consists of 60 interlaced fields per second[3]. The frame consists of 525 lines [21], of which the even and odd lines are alternated transmitted as fields.

24 Hz film Motion picture film is a system of recording moving images on a long strip of transparent material. The picture rate of 24 images per second is a compromise between the ability to capture motion and the amount of film material required per time interval. The standard is much older than the video transmission standards. Attempts were made to adapt the picture rate to 25 and 30 images per second, in order to become more compatible with transmission standards. Except for the recording of commercials, these frame rates did not find major ground in the motion picture industry. Therefore, 24 Hz film remains the most commonly used standard for motion pictures.

For a more detailed overview of the usage of these video standards, we refer to [21].

1.2 Standard Conversion Methods

When television became a popular medium, the need for new content increased. This called for format conversion methods. Besides converting motion pictures to television, television programs were converted between different transmission standards. Later, when the television became dominant, video material was converted to film, e.g. showing television commercials for cinemas.

In this report, we are interested in conversion methods from the film format to the video formats. Because of economic reasons, the motion picture industry still applies the traditional procedure of field repetition to transfer the film format to the video formats.

The process to transfer film to video is called the telecine process. One of the many implementations of this process is to illuminate the film and capture light coming though the film with a video camera and advancing the film in the vertical blanking period of the video signal.

To change the frame rate from 24 Hz film to 50 Hz video and 60 Hz video, a process called 'pull-down' is used. Pull-down is a sample and hold process, where the previous image from the film is repeated until a new one is available. This method can easily be implemented mechanically. The pull-down based telecine is the preferred method to transfer film to video. The general public accepts the motion artifacts introduced

[1] This is not entirely correct, since 50 Hz video was already used in black and white television transition and PAL and SECAM are a color encoding standards. Over time, both PAL and SECAM have become synonymous with 50Hz video.

[2] The remark about PAL is also valid for NTSC. NTSC is the color encoding standard used in regions where 60 Hz video is used. Over time, 60 Hz video and NTSC have become synonymous.

[3] 59.94 to be exact.

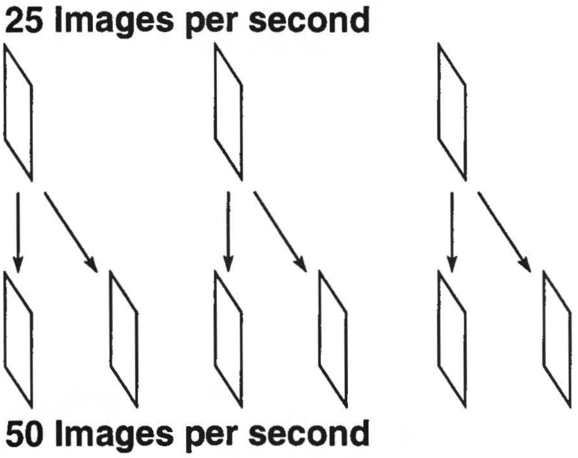


Figure 1.3: The 2:2 pull-down process. Each film frame is repeated, up converting 25 to 50 images per second.

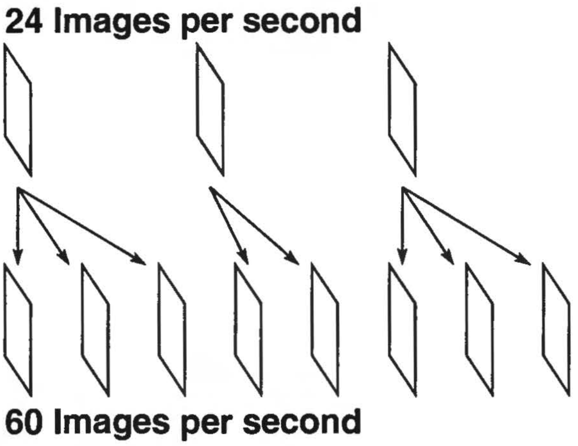


Figure 1.4: The 3:2 pull-down process. Each film frame is repeated three or two times. Up converting the 24 frames per second with a factor of 2.5, to 60 images per second.

by this method [17, page 110], as these motion artifacts are also present in the cinema. In cinemas the image is repeated multiple times to reduce large area flicker. The art director also minimizes these artifacts by employing techniques to reduce them e.g. using tracking shots (keeping the position of the camera the same in respect to the foreground object), long exposure times (by blurring motion the artifacts are reduced) and using small focus depth (the static foreground object is in focus, but the moving background is out of focus, hiding the artifacts).

To transfer 24 Hz film to 50 Hz video (figure 1.3), the frame rate of the film is increased to 25 frames per second by running the film 4% faster. This increase of the speed and the pitch of the sound is not regarded as annoying by the general public. Then, each film frame is scanned twice, creating two video fields. This method is called 2:2 pull-down.

To transfer 24 Hz film to 60 Hz video (figure 1.4), speed up to 30 Hz is not desired, since such a speed up and change in pitch of the sound is regarded as unacceptable by the general public. Therefore another method is used, where every even film frame is repeated three times while every odd film frame is repeated two times. This creates an increase of frame-rate by a factor 2.5, creating a video signal with a rate of 60 fields per second. This method is called 3:2 pull-down or 2:3 pull-down.

The sample and hold schemes used in these pull-down processes cause different fields, received by the video chain, to contain image data from the same temporal instance.

1.3 Film Detectors

Film detectors determine whether the pull-down process from 24 frames per second to 50 or 60 frames per second has been applied. This detection results in a classification of the video signal in either video mode, or film mode. This film mode can be characterized by the pull-down scheme used. The two schemes most common are 3:2 and 2:2 pull-down. Each scheme has its own sample and hold pattern of repeats of the image data.

In most solutions, the film detection is done in two stages. First the difference between two images is estimated. This is estimate is made by a difference metric. In the second stage a history of these difference estimates is analyzed. In this analysis, the temporal pattern detector determines the mode of the video signal. In the case of film mode, it also indicates which difference pattern is detected and at what position the current field is in that pattern.

Common metrics for 3:2 pull-down take the absolute difference between the current field and the pre-previous field. Such a measurement does not suffer from the interlacing process, since both the fields have the same interlace phase, i.e. they both have either an even phase or have an odd phase.

In metrics for 2:2 pull-down, an absolute difference between current field and the previous field is taken. These two fields always have different interlace phases. To compare these fields, one of the fields has to be de-interlaced. De-interlacing without prior knowledge of the film mode can introduce artifacts in the de-interlaced picture. These artifacts introduce differences that are interpreted as motion by the film detector. This has a negative impact on the robustness of the film detection.

The challenge in film detection lies in robustly detecting 2:2 pull-down film mode, while minimizing the resources required by such a detector.

1.4 This project

Philips has been researching film detectors since the beginning of the 1990s. In their 100 Hz television sets, they employ film-detection to correctly de-interlace video signals. These film detectors were implemented in the SAA7158 (Bendic), SAA4990 (Prozonic) [34], SAA4991 (Melzonic) and SAA4993 (Falcon) [35] video processing ICs.

The Bendic and Prozonic implement non-motion compensated de-interlacing. An accumulated absolute difference taken directly between two fields is used to detect whether the incoming video signal is in film mode. This difference taken directly between two fields ignores the fact that the video signal is interlaced. This reduces the complexity of the implementation, but results in a lower reliability than comparing the fields using a de-interlace scheme.

The Melzonic and Falcon image processors use motion compensated de-interlacing to eliminate motion judder created during the pull-down process. The film detector in these video processors is based on motion vectors [18] instead of direct differences between successive images, the method used by the Bendic and the Prozonic processors. The rationale behind this design choice is that the motion vectors are already available in the video processing chain and that these motion vectors suffer less from the interlacing of the video signal.

A new video processing IC was proposed to succeed the Falcon. The new IC is the Condor video processor. This processing system contains a stand-alone film detector, the *Schutten Riemens Film Detector* (SRFD).

In this project, we focus on this SRFD and the implementation of the SRFD on the Nexperia platform [4]. The Nexperia platform uses the Tri-Media video processing IC [38, 37]. An application of this platform is a scan-rate converter used in consumer electronics, e.g. high-end television sets. In this application field, the cost of the implementation is important. The cost will determine the validity of the film detector as a viable system. This criterion is as important as the performance.

Taking these conditions into account, we formulate two goals for this project:

- Gaining an overview of the field of film detection technology.

[4] For more information, see: *http://www.semiconductors.philips.com/platforms/nexperia*

- Improving the performance of the SRFD without significant increase in algorithmic cost.

We achieve the first goal by doing an extensive literature research, which is reported in Chapter 3. Using this knowledge we improve the performance of the SRFD.

We improve the performance of the SRFD by analyzing the behavior of the SRFD on a number of typical sequences. We have collected a number of sequences that are detected incorrectly by the SRFD. By analyzing these problem sequences, we have identified several problems.

One of the problems that is identified is vertically detailed areas that are interpreted as motion. In Chapter 5, we propose and test a new motion detection scheme, the Arrow detector. The Arrow detector is designed to reduce incorrect film mode detections caused by static vertically detailed areas.

This solution was implemented on, and optimized for, the Tri-Media processor. In this way, a realistic comparison with the original metric of the SRFD can be made in terms of algorithmic cost. This report does not cover the process of optimizing the implementation in detail, only the reduction in cost of this optimization.

Additionally in Chapter 5, we implement and test a hard-cut shot boundary detector. A shot is a sequence of images recorded without interruption. A hard-cut shot boundary is an abrupt transition from one shot to another. Although this shot boundary detector does not directly contribute to the performance of the film detector, we hope to use it in the future. The shot detector can be used indirectly to detect an artifact called video edits, a problem identified in the literature research.

In Chapter 6, the SRFD with the Arrow difference metric is evaluated and conclusions drawn from this evaluation.

Chapter 2

Terms and Definitions

In this chapter we introduce some terms and definitions that we will commonly use throughout this report.

2.1 Video Signals

We use in this report, a spatial-temporal sampled version of a luminance video signal: $F[n]$. The signal $F[n]$ is an array of fields, received in an interlaced fashion, where n indicates the field number. The field number indicates the order of reception of the image. Since the film detector has to determine, whether a sampling and hold scheme has been applied to this array, it does not indicate the moment that the field $F[n]$ is sampled. The value n merely indicates the order of the fields.

Each field consist of a two dimensional (spatial) array of luminance values $F(\vec{x}, n)$ [2]:

$$F(\vec{x}, n) = \begin{cases} F_{\text{original}}(\vec{x}, n) & \forall y \bmod 2 = n \bmod 2 \\ \text{undefined} & \text{otherwise} \end{cases} \tag{2.1}$$

with $\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ designating the discrete (integer) spatial position, n the discrete (integer) temporal position and $F_{\text{original}}(\vec{x}, n)$ is the luminance signal registered by the camera.

We also define a set $A(n)$, containing all or a subset of the spatial positions in $F[n]$.

2.2 Pull-down

The term pull-down originates from the process used in telecine devices. Traditionally, in a telecine device, a film frame is scanned by a video camera. Nowadays, other means are used, but the process is basically the same. The image was scanned multiple times to compensate for the difference in frame rate. After a sufficient number of scans were made, the filmstrip was pulled down, advancing the film to the next frame. This created a pattern of a number of repeated scans. Nowadays, the term pull-down is used for all video material that has repeated images. By determining the moments where the luminance data was originally registered, the film detector can determine whether a sampling and hold scheme has been applied to the fields. If a film frame is transferred to a video field, and then transferred to another video, both video fields contain data from the same registration moment.

We define two kinds of motion behavior:

Video motion This type of motion has not been affected by the pull-down process. This kind of motion characterizes sequences in video mode.

Pull-down motion This type of motion has been affected by the pull-down process. This kind of motion characterizes film mode.

2.3 Difference Numbers

We define the output of a difference value as δ. We define $\delta[n]$ as the difference between field n and a previous version. Examples of such difference metrics are the field difference $\delta_{\text{field}}[n]$ and the frame difference $\delta_{\text{frame}}[n]$.

Film detectors attempt to detect these repeated images by comparing fields. If two fields are scanned from the same film frame, they should be similar. However, due to various processes applied after conversion to video (e.g. interlace, noise), the difference between the images usually is not equal to zero. Therefore we use the `HL` classification to describe the difference patterns. We assign a classification to the variable $\delta[n]$. The two values that are assigned are:

High The difference value $\delta[n]$ is high, indicating that the two fields are not the same. A high value is indicated by the symbol `H`.

Low The difference value $\delta[n]$ is low, indicating that the two fields are the same. A low value is indicated by the symbol `L`.

Using this classification, a temporal difference pattern can look like:
`HLHLHLHL`
which is shorthand for:

$$\begin{aligned} \delta[n] &= \mathtt{L} \\ \delta[n-1] &= \mathtt{H} \\ \delta[n-2] &= \mathtt{L} \\ \delta[n-3] &= \mathtt{H} \\ \delta[n-4] &= \mathtt{L} \\ \delta[n-5] &= \mathtt{H} \\ \delta[n-6] &= \mathtt{L} \\ \delta[n-7] &= \mathtt{H} \end{aligned}$$

A `H` indicates that two images originate from the same temporal position, i.e. they are either scanned by a video camera, or they originate from different film frames. A `L` indicates that two images either originate from the same temporal position, or that there is no motion in the sequence.

The value of $\delta[n]$ is often based on a spatial luminance difference $\delta(\vec{x}, n)$. This spatial luminance difference compares luminance values in two or more fields at (about) the same spatial position $\vec{x}$. An example of such a spatial luminance difference is the spatial frame difference:

$$\delta_{\text{frame}}(\vec{x}, n) = |F(\vec{x}, n) - F(\vec{x}, n-2)| \tag{2.2}$$

In order to calculate a difference number, we use the # operator:

$$N = \mathop{\#}_{a \in A} a \tag{2.3}$$

where A is a (discrete) set of Boolean values and N is the number of elements in A that equal true.

2.4 Film Detector Output

The output of a film detector is the:

mode This indicates whether a video signal is in video mode, or in film mode. If the video signal is in film mode, it also indicates what kind of pull-down scheme is used. In this report, this signal is indicated with the symbol M. M can take the values *video mode*, *2:2 pull-down film mode* and *3:2 pull-down film mode*.

phase This indicates the phase of the film mode, the current position in the difference pattern. This is indicated with the symbol ϕ. Based on the value of M, the number of phases available ϕ depends on the length of the repeat pattern for that mode. The number of phases are shown in table 2.1.

Table 2.1: The modes and the number of phases per mode and examples of the available phases.

M	number of phases ϕ	example
video mode	1	HHHHHHHHHH
2:2 pull-down film mode	2	HLHLHLHLHL
		LHLHLHLHLH
3:2 pull-down film mode	5	HLHLLHLHLL
		LHLLHLHLLH
		HLLHLHLLHL
		LLHLHLLHLH
		LHLHLLHLHL

Using these two signals, the scan-rate converter can adapt its video processing scheme based on M. Using the phase ϕ, the scan-rate converter can determine how the fields should be combined to correctly de-interlace the signal.

Chapter 3

State of technology

This chapter covers the search methods on literature covering film detector technology we have used and an overview of the prior art of film detector technology based on the results of that literature research.

In section 3.1, the method used in the literature search is described. Next, in section 3.2 the prior art on the film detection technology is discussed. In sections 3.3 and onward, an overview of the literature is presented. Finally in section 3.9, the conclusions drawn in this chapter are summarized.

3.1 Literature research

The search was conducted into two iterative steps. The first part was conducted using 'Bakker's method'. The second iteration used only the keyword search on the PASS patent database. This keyword search on the PASS patent database was based on the result of the first iteration.

I used Bakker's method for the initial literature research. Using this method is a mandatory part of my final project at the Eindhoven University of Technology. The search method is prescribed by the Eindhoven University of Technology's Electrotechnical department's librarian, ir. R.G. Bakker. The results of this search method is reported in Appendix A.

The method consists of several steps:

Keyword Search Key documents are located by searching specific databases. In this case, these include the Eindhoven University of Technology's library and the INSPEC reference database.

Snowball Method Using the key documents, the snowball method is started. This method uses the (relevant) references of a document. With these references, new references are identified. In this way, the history of a document can be constructed.

Citation Method Using the key documents, the citation method is started. This method looks for the documents that cite the key document. The documents that cite the current document are used for the next iteration step, creating an overview of the publications that relate to the key document.

If necessary, the search method can be repeated until a sufficient number of documents have been found. The report of the literature search is included as Appendix A.

Additional to this method, in the second iteration, a keyword search was performed on the patent library. The results of this search are incorporated in this report. The keywords used are 'film', 'video', 'pull' and 'down'. The choice of these keywords is based on the papers that resulted from Bakker's method. The results of this search are included as Appendix B.

3.2 Prior art of film detection technology

Film detector technology evolved from other technologies. Some of these technologies, like transmission standards and standard conversion are considered trivial in the field of video processing. In order to get a better view on the origin of film detector technology, the result from the snowball method of the literature research is used to determine the technologies preceding film detector technology. In section 3.3, the origin of film detectors is examined using the results of the snowball method. The conclusions draw from these sections are summarized in section 3.2.2.

3.2.1 Technologies preceding film detectors

The overview in figure A.1 was created by checking the references of the paper by Armitano [1] about film detection. By continuously checking references until we end at a 'root' document, we can follow the references to see the origins of film detector. Almost all publications in this diagram are patents. This can be explained by the fact that the research into this area is done at industrial research institutes and the fact that film-detection is only part of a larger problem and may depend on the application.

Several trends are observed in the results of the snowball method. These trends are:

- Film detector patents appeared from about 1990.
- Film detector technology originates from de-interlacer and telecine devices.
- Frame rate conversion and de-interlacer patents appeared from the mid 1980s.
- Prior Art before the mid 1980s consists of telecine devices.

The de-interlacer devices and schemes [45, 31, 12, 15] and frame rate converters [36, 47] show a trend in improving the image quality for film material. These schemes are incorporated into film phase detector devices [13, 23] that can correctly de-interlace video signals in film mode. These devices cannot detect the mode of the video signal, but assume that the incoming video signal is in film mode. This excludes those film phase detectors as being classified as film detectors.

Based on these film phase detection systems, de-interlacers and frame rate converters, systems were devised that could determine the mode of the video signal. The ability to distinguish different video modes classifies those systems as film detectors.

3.2.2 Conclusions

From this section the following conclusions can be drawn:

- Except for Armitano [1], all publications on this subject are patents or patent applications. This suggest that this problem field is highly specialized and is only covered by industrial research.
- The art known before film detection consists of video and film standards and frame rate conversion methods. Telecine devices, that transfer film material to a video signal, use these methods. The method is a sample and hold scheme, also known as pull-down.
- Based on the trends observed in the snow ball method, it can be observed that film detector technology seems to originate from telecine devices, frame rate converters, de-interlacers and film phase detectors.
- The first film detector patents appeared around 1990. The earliest patent found during this research is by Lyon and Campbell [28].

3.3 Overview of film detector technology

In this section, we summarize the common elements in the trends in film detection. This way, we give an overview of the current state of technology. For the analysis we use the publications found with the search method described in section 3. In the next section, elements common to film detectors are presented. Next, in section 3.3.2, we discuss the trends in the field. Finally, in section 3.9, we conclude this chapter by summarizing the common elements and trends and show how these trends are related.

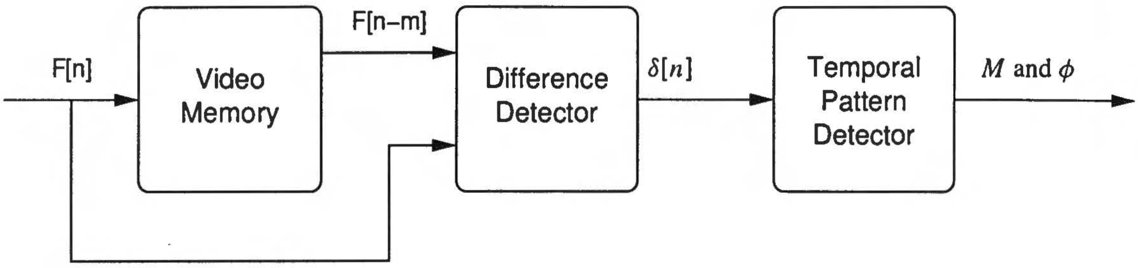


Figure 3.1: The general structure of a film detector. The difference detector takes the current video signal $F[n]$ and a m field delayed version $F[n-m]$ and calculates motion number $\delta[n]$ for the current field. The field motion number $\delta[n]$ indicates the amount of motion in the current field and is then stored in the temporal pattern detector to detect the film mode M and film phase ϕ of the incoming video signal $F[n]$.

3.3.1 Common elements

Within the reviewed patents, the following common elements can be observed. These common elements are the input, an interlaced luminance signal, the output, a set of discrete signals indicating the mode and phase of the input and a general structure.

Input

All the film detectors reviewed use an interlaced luminance signal $F[n]$ as input.

Output

Within the area of film detection, three modes are used. Video mode, 2:2 pull-down film mode and 3:2 pull-down film mode.

Both the telecine process for 50 Hz video and from progressive scan cameras produce 2:2 pull-down film mode for both 50 Hz and 60 Hz video, so 2:2 pull-down is not limited to 50 Hz. The occurrence of 3:2 pull-down however, is, according to the literature, limited to 60 Hz video.

The output is generally film-mode signal mode M and a film-phase signal ϕ.

General structure

In general, film detectors can be described as a three-part structure. This structure is shown in figure 3.1.

Video Memory A video memory produces a time delayed version of a video signal. Three typical sizes of video memories are line memories, field memories and frame memories. Video memories can be cascaded to construct a video memory with a longer delay. Using these memories, time-delayed versions of the video signal $F[n]$ are generated. The two most common are the field period delayed version $F[n-1]$ and the two field-period or frame-period delayed version $F[n-2]$.

Difference Detector A difference detector is a device that detects whether two fields originate from the same film frame. These detectors can be classified into four groups, based on the approach they used:

- Detectors that try to match the zero motion vector on a previous field.
- Detectors that try to detect horizontal jagged edges in the frame.
- Edge detection based detectors.
- Motion estimation based detectors.

The detector uses the luminance signal $F[n]$ and one or more time delayed versions. The output of the difference detector is an estimate of the amount of motion $\delta[n]$ that indicates whether there was motion or the amount of motion in $F[n]$.

Temporal Pattern Detector The temporal pattern detector takes the difference value $\delta[n]$ and extracts information about the film mode of the incoming video signal. To this end, temporal patterns in the estimated amount of motion signal produced by the difference detector are analyzed. The output is a set of discrete values indicating the film mode M and film phase ϕ of the video signal.

This structure for film detectors uses a two-step feature extraction. First the amount of motion for the entire frame is extracted as a motion signal. From these motion signals, the temporal pattern detector determines the mode of the video signal. This structure is claimed by Faroudja Laboratories in a patent by Lyon and Campbell [28].

3.3.2 Film detector properties

We analyze the trends in the field of film detectors, by dividing the field of film detection based on their properties. These properties are based on key differences in components and signals and on the groups of solutions proposed to deal with the problems found in film detection. For each of the properties, a categorization is made.

Type of difference pattern used Two types of temporal difference patterns are identified. The type of difference pattern defines the ability to detect certain types of film modes and their robustness of that detection.

Type of difference detector used Several approaches have been proposed to detect differences between fields. These approaches are zero vector matchers, jagged edge detection, edge based detection and motion vector based detection.

Field of application The field of usage of film detectors. These fields are scan-rate conversion (e.g. de-interlacing) and video encoding for compression (e.g. MPEG-2 encoding).

Of these classifications, the first, based on the type of difference pattern used, is the most important one. The type of difference pattern determines the capabilities and is a major influence on the quality of the film detector.

Besides these classifications, several problems with film detection are identified. The most important are false detection due to vertical detail, video edits, false detection due to noise in the input signal and mixed mode due to overlays. Several solutions for these can be categorized.

Reducing false detections due to static vertical detail Several solutions have been proposed to reduce the effect of false motion detection due to static vertical detail.

Zoning In order to increase the robustness of the film-detectors, zoning strategies have been proposed.

Classifications not described in this report are those based on detection of video edits and based on reducing sensitivity to noise.

3.4 Types of temporal difference patterns

This section discusses the temporal difference patterns that are mentioned in the prior art. The temporal difference patterns are an important property of the film-detector. The type that is used determines to a large extent the capability of the film detector. Each type consists of a set op patterns. Each pattern can be associated with a film mode. The types of temporal difference patterns that are present in the prior-art are:

Frame difference patterns These patterns (table 3.1) is generated by the difference detector measuring between the current field n and the pre-previous field $(n-2)$. Video mode and 2:2 pull-down film mode both have the same pattern. A film-detector using this type cannot distinguish between video mode and 2:2 pull-down.

This type of difference patterns is used by Casavant et al. of Thomson [4], Yonemitsu et al of Sony [48], Lee et al of Electronics and Telecommunications Research Institute, Daejeon, Korea [24],

mode	difference pattern
video	HHHHHHHHHHHHHHH
2:2 pull-down film	HHHHHHHHHHHHHHH
3:2 pull-down film	HHHHLHHHHLHHHHL

Table 3.1: Typical motion patterns produced by a difference detector measuring over a frame period.

mode	difference pattern
video	HHHHHHHHHHHHHHH
2:2 pull-down	HLHLHLHLHLHLHLH
3:2 pull-down	HLHLLHLHLLHLHLL

Table 3.2: Typical motion patterns produced by a difference detector measuring over a field period.

Yagasaki et al of Sony [46], Lim of LG Electronics [25] and Del Corso of Royal Philips Electronics [11].

Field difference patterns These patterns (table 3.2) are generated by a difference detector measuring between the current field n and the previous field $n-1$. Because each of the three modes produces a distinct motion pattern, the temporal pattern detector can detect the mode of the incoming video signal as 'video', '2:2 pull-down film mode' or '3:2 pull-down film mode'.

This type of difference patterns is used by Correa and Schweer of Deutsche Thomson-Brandt GmbH [9, 10], Christopher and Correa of Thomson Consumer Electronics [5, 6, 7], Swartz of Faroudja Laboratories [42, 43], Gerets of Barco N.V. [16], Swan of ATI Technologies [40], and Faroudja et al of Faroudja Laboratories [14] and Coombs et al of Philips Electronics UK Limited [8].

Frame difference patterns do not suffer from the interlacing of the video signal. They can be detected more robustly than field difference patterns because of this. However, they have the drawback that they cannot be used to distinguish 2:2 pull-down film mode from signals in video mode. Also, a field-difference pattern based film detector can detect the mode of the video signal faster, because the patterns has more feature points, when compared to frame difference patterns.

To take advantage of both types of patterns, hybrid solutions exist that use both types of difference patterns. These solutions have the robustness of the 'frame difference patterns' and the resolution and speed of the 'field difference patterns'. By running two difference detectors in parallel, these two patterns are generated. Solutions using this scheme are proposed by Lyon and Campbell of Faroudja Laboratories [28] and Hui of STMICRO-Electronics Asia Pacific PTE LTD [26].

3.5 Types of difference detectors

In this section, we give an overview of the methods used generate the difference value $\delta[n]$. In the publications on film detection, four approaches are identified.

Zero vector matching This class of detectors tries to match the zero motion vector directly on a previous field. The absolute difference of two luminance values is taken to determine whether the zero vector matches. A mis-match over the zero vector in the motion estimation indicates that there is motion. This is used to generate the difference value $\delta[n]$.

Horizontal jagged edge detection When two fields from different temporal positions are combined into a frame, jagged horizontal edges can be observed. Jagged edges are spatial-temporal property caused by motion. This class of difference detectors uses this property to generate the difference number $\delta[n]$.

Edge detection based This class of detectors attempts to detect if fields originate from different temporal positions by the edges in a sequence. These edges are used to generate a difference value $\delta[n]$.

Motion vector based The summed length of the vector field that has been generated by a motion estimator. This vector field is used to generate a difference value $\delta[n]$.

Each of these approaches will be described in more detail in the next sections.

3.5.1 Zero vector matching

The idea behind the zero vector matching approach is to detect whether the zero vector matches between the luminance values of two fields, at the same spatial position. A match is made if the absolute difference between two luminance values is smaller than a preset threshold. Several methods are proposed to measure this: zero vector matching on a frame difference, zero vector matching on a field difference and zero vector matching on a field difference using a de-interlacer.

This subsection describes a kernel of four luminance values P_1, P_2, P_3 and P_4 (figure 3.2a). P_1 is the 'current' value $F(\vec{x}, n)$. P_2 is the vertical neighboring value in the previous field $F(\vec{x} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}, n-1)$. P_3 is the other vertical neighboring value in the previous field $F(\vec{x} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}, n-1)$. P_4 is the value at the same spatial position as P_1 in the pre-previous field, $F(\vec{x}, n-2)$

Two methods are discussed to generate the difference value $\delta[n]$:

- By counting the number of absolute differences that exceed a threshold:

$$\delta[n] = \underset{\vec{x} \in A(n)}{\#} (\delta(\vec{x}, n) > \text{Thr}) \qquad (3.1)$$

  Here the value $\delta[n]$ is an integer value indicating the amount of motion. The set $A(n)$ contains all spatial positions $\vec{x}$ checked for field n.

- By applying a threshold to the sum of the absolute differences

$$\delta[n] = \left(\sum_{\vec{x} \in A(n)} \delta(\vec{x}, n) \right) > \text{Thr} \qquad (3.2)$$

  Here the value $\delta[n]$ is a Boolean value indicating the presence of motion. The set $A(n)$ contains all spatial positions $\vec{x}$ checked for field n.

Zero vector matching on a frame difference

This approach detects if the zero vector matches between a field n and the pre-previous field $n-2$ (figure 3.2b). The absolute luminance difference between two frames is calculated. For a single spatial position $\vec{x}$ this absolute difference $\delta_{\text{frame}}(\vec{x}, n)$ is calculated as:

$$\delta_{\text{frame}}(\vec{x}, n) = |F(\vec{x}, n) - F(\vec{x}, n-2)| \qquad (3.3)$$

When the fields n and $n-2$ originate from the same film frame, the absolute difference $\delta_{\text{frame}}(\vec{x}, n)$ equals zero. To determine the amount of motion in the entire field, the absolute difference is accumulated over the set $A(n)$. $A(n)$ contains a set of luminance values of the field n:

$$\delta_{\text{frame}}[n] = \sum_{\vec{x} \in A(n)} \delta_{\text{frame}}(\vec{x}, n) \qquad (3.4)$$

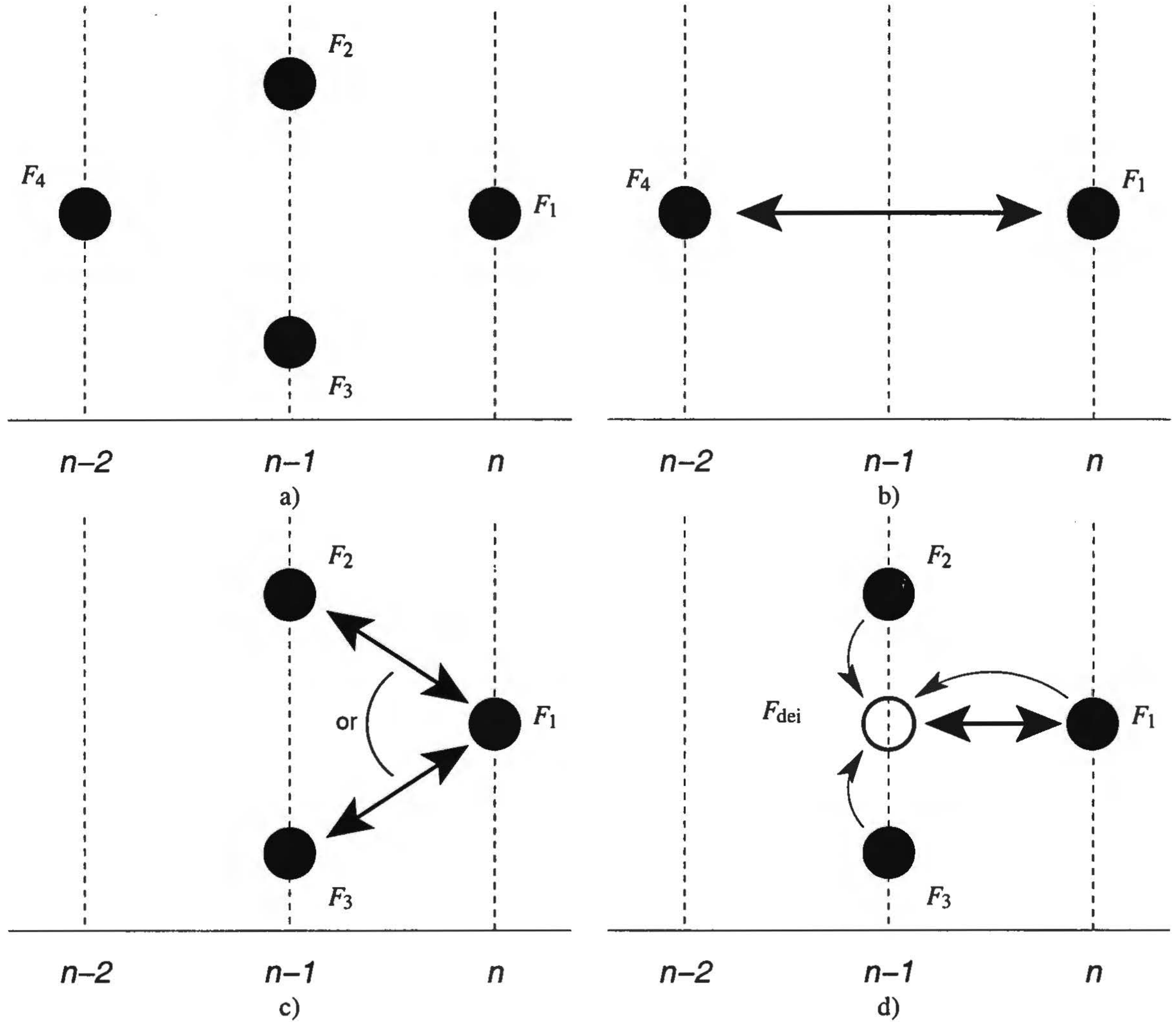


Figure 3.2: The different methods for used as zero vector matching. *a)* is the general support used by all methods. The support consists of four luminance values, F_1 being the value currently processed. F_2 and F_3 are the two vertical neighbors in the previous field $n-1$ and F_4 is the value at the same position as F_1, but in the pre-previous field $n-2$. *b)* The comparison of between the current value F_1 and the value in pre-previous field F_4. *c)* The comparison between the current value F_1 and one of the spatial neighbors in the previous field F_2 or F_3. *d)* The comparison between the current value F_1 and a de-interlaced value F_{dei}. The de-interlaced value is calculated by taking the median of the values F_1, F_2 and F_3.

Noise added during transmission can create differences between the two fields, creating a non-zero value for $\delta_{\text{frame}}(\vec{x}, n)$. To prevent false motion detection, due to noise, a coring or threshold step is introduced. The position of this step depends on the proposed implementation. It has been proposed before as well as after the accumulation step.

This method of pull-down detection is used by Armitano [1], Lyon and Campbell of Faroudja Laboratories[28], Casavant et al. of RCA Thomson [4], Jun Yonemitsu et al. of Sony [48], Bock of Digi-Media Vision [3], Lee et al. of the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea [24], Yagasaki and Suzuki of Sony [46], Hui of STMICRO-Electronics Asia Pacific PTE [26], Lim of LG Electronics [25] and Del Corso of Royal Philips Electronics [11].

Zero vector matching on a field difference

Analog to zero vector matching on a frame difference, this approach compares the a luminance value with a luminance value in the previous field at (about) the same spatial position (figure 3.2c). In this case the difference between the current field n and the previous field $n-1$:

$$\delta_{\text{field}}(\vec{x}, n) = \left| F(\vec{x}, n) - F(\vec{x} \pm \begin{pmatrix} 0 \\ 1 \end{pmatrix}, n-1) \right| \tag{3.5}$$

Depending on the implementation, the detector compares the luminance value $(\vec{x}, n)$ with the value in the previous field above the current value $(F(\vec{x}, n) - F(\vec{x} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}, n-1)$ or below the current value $((F(\vec{x}, n) - F(\vec{x} - \begin{pmatrix} 0 \\ 1 \end{pmatrix}, n-1)$.

To determine the amount of motion in the entire field, the absolute difference is calculated over the set $A(n)$. $A(n)$ contains a set of luminance values in the field n.

$$\delta_{\text{field}}[n] = \sum_{\vec{x} \in A(n)} \delta_{\text{field}}(\vec{x}, n) \tag{3.6}$$

This method is used in Philips' SAA7158 (Bendic) and SAA4990 (Prozonic) [34] video processing ICs, by Lyon and Campbell of Faroudja Laboratories [28] and by Coombs et al. of Philips Electronics UK [8].

Zero vector matching using a de-interlacer

This method is a refinement on the 'direct comparison using a field delay'. The current field n is compared with a de-interlaced version of the previous field $n-1$ (figure 3.2d). By de-interlacing the previous field, the zero vector between the current field and that field can be tested.

$$\delta_{dei}(\vec{x}, n) = |F(\vec{x}, n) - F_{dei}(\vec{x}, n-1)| \tag{3.7}$$

The de-interlacer proposed by Christopher and Correa of Thomson Consumer Electronics [5, 6, 7] and Gerets of Barco N.V. [16], is a three-tap VT median de-interlacer (equation 3.8). This median de-interlacer uses a kernel of the the present luminance value $F(\vec{x}, n)$ and the two vertical neighbors $F(\vec{x} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}, n-1)$ and $F(\vec{x} - \begin{pmatrix} 1 \\ 0 \end{pmatrix}, n-1)$ in the previous field.

$$F_{\text{dei}}(\vec{x}, n-1) = \text{med}\left(F(\vec{x}, n), F(\vec{x} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}, n-1), F(\vec{x} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}, n-1) \right) \tag{3.8}$$

The function med(A, B, C) is defined by:

$$\text{med}(A, B, C) = \begin{cases} A & (B < A < C) \vee (C < A < B) \\ B & (A \leq B \leq C) \vee (C \leq B \leq A) \\ C & \text{otherwise} \end{cases} \quad (3.9)$$

Discussion

Each type of zero vector matching has its own advantage.

Zero vector matching on a field difference This method is the most economic version, since it uses no de-interlacer and only a single field memory. The method's performance as a pull-down detector is poor, since the method does not test whether the zero vector, but whether the vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ (or $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$) matches between the two fields. The method ignores the fact that the video signal is interlaced. This causes false movement detection on horizontal edges and areas containing static vertical detail. The method is able to detect both 3:2 and 2:2 pull-down.

Zero vector matching using a de-interlacer This method is the most universal solution. It can be used to detect both 3:2 and 2:2 pull-down patterns. This does not suffer from horizontal edges, like the 'zero vector matching on a field difference' but still suffers from false movement detection caused by alias. The three-tap VT median algorithm distorts vertical detail and introduces alias [17, page 161]. This alias is detected as false motion, potentially causing faulty detections of the film mode.

Zero vector matching on a frame difference This method can only be used to detect 3:2 pull-down patterns. The detector is better in detecting 3:2 pull-down than the other two methods, because it does not have to deal with interlacing, since the two fields that are compared always have the same interlace phase. A disadvantage is, that it cannot detect 2:2 pull-down patterns. Another disadvantage is that it requires a delay that is twice the delay used in the other two zero vector matching methods.

Zero vector matches all implement a threshold or coring step to reduce the effect of noise. Noise causes differences between the fields that is detected as motion. The threshold or coring step can remove this effect, but forces a trade-off between noise insensitivity and movement sensitivity, e.g. the higher the threshold, the lower the sensitivity to of noise, but also the lower the sensitivity to motion.

3.5.2 Jagged edges detection

Another approach for solving the problem of pull-down detection is to detect jagged edges in frames. These edges are high frequency vertical transitions at half the vertical sampling frequency. They occur when two fields, sampled at different moments in time, with moving objects are merged into one frame. Due to the fact that the objects are present at different positions in both of the fields, jagged edges or jaggies appear.

Figure 3.3 illustrates this. If we take a film sequence with a moving object, and combine two fields from different film frames, field 3.3 a) and field 3.3 b) into frame 3.3 c), jagged edges will appear. These edges occur, because frames a) and b) are sampled at different moments in time. The jagged edges are an indication that the two fields in a frame are sampled at different moments in time and thus originate from different film frames. The absence of these jagged edges can indicate that the two fields originate from the same film frame.

In a video signal that originates from film, the occurrence of jagged edges forms a temporal pattern. This pattern is similar to the pattern found by the zero vector matchers. In video mode, the jagged edges occur in every frame, because each field originates from a unique point in time.

Several variants of this kind of detector have been proposed. The ones we will illustrate are the proposals by Hui [26] and Correa and Schweer [9, 10].

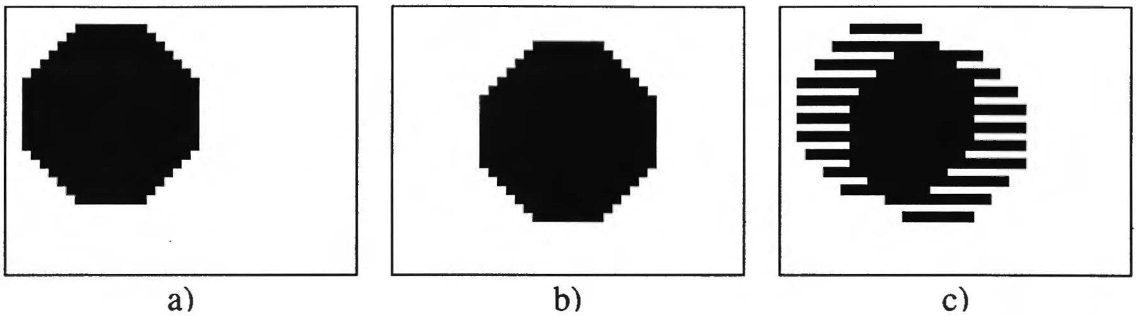


Figure 3.3: Frame c) contains a field from frame a) and a field from frame b).

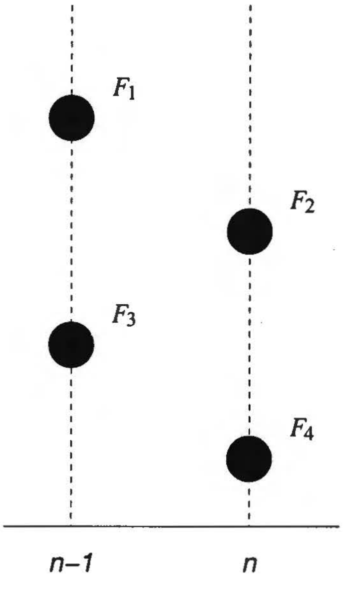


Figure 3.4: The support for the jagged edge detector proposed by Hui.

Jagged edge detection by Hui

A Jagged edge detector is proposed by Hui from STMICRO-Electronics Asia Pacific PTE [26]. The detector takes differences between luminance values in the same column from different fields. The support (equation 3.11) for this detector consists of the luminance values F_1, F_2, F_3 and F_4, as shown in figure 3.4). Using this support, Hui count the number of times that the differences between the two fields exceeds the threshold Thr. In the presence of jagged edges, the differences can exceed the threshold.

$$\begin{aligned}\delta_{\mathrm{hui}}[n] = & \underset{\vec{x}\in A(n)}{\#}\big((F_1 - F_2) > \mathrm{Thr} \wedge (F_3 - F_2) > \mathrm{Thr} \wedge (F_3 - F_4) > \mathrm{Thr}\big) \\ & + \underset{\vec{x}\in A(n)}{\#}\big((F_1 - F_2) < \mathrm{Thr} \wedge (F_3 - F_2) < \mathrm{Thr} \wedge (F_3 - F_4) < \mathrm{Thr}\big)\end{aligned} \tag{3.10}$$

with

$$
\begin{aligned}
F_1 &= F(\vec{x} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}, n-1) \\
F_2 &= F(\vec{x}, n) \\
F_3 &= F(\vec{x} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}, n-1) \\
F_4 &= F(\vec{x} + \begin{pmatrix} 0 \\ 2 \end{pmatrix}, n)
\end{aligned}
\tag{3.11}
$$

Jagged edge detection by Correa and Schweer

The method proposed by Correa and Schweer of Deutsche Thomson-Brandt GmbH [9, 10] tries to detect these edges by applying equation 3.12. In equation 3.12, $\delta_{CS}[n]$ is the difference value for field n. The sign of indicator of shows the phase of the pull-down. If we use the support shown in figure 3.5, $\delta_{CS}[n]$ is calculated as:

$$
\begin{aligned}
\delta_{CS}[n] = &\underset{\vec{x}\in A(n)}{\#}\Big(\big(F_1 < \min(F_2, F_3) \vee F_1 > \max(F_2, F_3)\big) \wedge \big(\min(F_2, F_3) \leq F_4 \leq \max(F_2, F_3)\big)\Big) \\
&- \underset{\vec{x}\in A(n)}{\#}\Big(\big(F_4 < \min(F_2, F_3) \vee F_4 > \max(F_2, F_3)\big) \wedge \big(\min(F_2, F_3) \leq F_1 \leq \max(F_2, F_3)\big)\Big)
\end{aligned}
\tag{3.12}
$$

with

$$
\begin{aligned}
F_1 &= F(\vec{x}, n) \\
F_2 &= F(\vec{x} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}, n-1) \\
F_3 &= F(\vec{x} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}, n-1) \\
F_4 &= F(\vec{x}, n-2)
\end{aligned}
\tag{3.13}
$$

Here F_1, F_2, F_3 and F_4 (equations 3.13) are luminance values. The difference value $\delta_{CS}[n]$ is incremented if for the current spatial position $\vec{x}$ there is a jagged edge detected between fields $F[n]$ and $F[n-1]$ and no jagged edge is detected between fields $F[n-1]$ and $F[n-2]$. The value of $\delta_{CS}[n]$ is decremented if there is a no jagged edged detected between fields $F[n]$ and $F[n-1]$ and there is a jagged edge detected between fields $F[n-1]$ and $F[n-2]$. A similar implementation of this strategy is used by Christopher and Correa from Thomson [5, 6, 7].

Other proposals

Other proposals were made by Swartz of Faroudja Laboratories [41, 42, 43] and Faroudja et al. of Faroudja Laboratories [14]. They employ a magnitude comparison to detect jagged edges.

3.5.3 Edge based detection

An edge based detection scheme is proposed by Swan of ATI Technologies[40]. Horizontal edge positions are encoded into signatures. These signatures are then compared to detect film motion patterns. If two fields originate from the same film frame, the edges in those images should be at the same spatial position. These identical positions should result in a similar signature. These similarities in signatures are used to produce a difference value $\delta[n]$.

3.5.4 Motion estimation based

A motion vector based approach is proposed by de Haan et al [19]. Alternating patterns of motion and no-motion can be observed in the sum of the length of the motion vectors. This sum is used to produce a film mode and phase of the sequence.

3.5.5 Discussion

Hui's approach is the most direct way to detect jagged edges. This detector simply checks if the differences between successive lines in a frame exceed a threshold. It probably will detect false motion in areas containing high vertical detail. To counter this, the method also implements zero vector matching on a frame difference and combines the output of the two detectors to increase the robustness.

The jagged edge detector proposed by Correa and Schweer can only detect motion that has been processed by the telecine process. This is discussed in more detail in section 3.7.1. The detector is blind for any other kind of motion, i.e. motion in video mode. Therefore, long-term temporal disturbances with a small intensity can cause an incorrect detection. A threshold to eliminate the effect of noise is not implemented. The assumption that noise is randomly distributed (spatial and temporal) is used. This random distribution causes random values of 1 and -1 to be added to a, canceling out the effect of noise.

The proposal by Swan of ATI Technologies [40] uses less memory than the other proposals in this chapter. It only needs to store one signature per line. The performance of this edge based film detector is unknown to us at this moment.

Zero vector matching, jagged edge detection and motion vector based approaches are proven methods to detect motion patterns in video sequences for the purpose of film detection.

The large number of jagged edge and zero-vector matching patterns, when compared to the other methods, indicate that this kind of approach is the most popular.

3.6 Application areas

In this section, we try to give an overview of the application field of film detectors. Based on the patents found in the literature research, the field can be divided into two sections:

Frame-rate conversion Film detection is used in frame-rate conversion. An example is de-interlacing. Perfect de-interlacing can be achieved by merging the fields into their original film frames. This technique is employed as a part of flicker reduction in high-end televisions. Lyon and Campbell of Faroudja Laboratories [28] proposal is a good example of film-detector designed to control a de-interlacer. Other proposed implementations for de-interlacers are from Correa and Schweer of Deutsche Thomson-Brandt GmbH [9, 10], from Christopher and Correa of Thomson Consumer Electronics [5, 6, 7], from Swartz of Faroudja Laboratories [41, 42, 43], from Gerets of Barco N.V. [16], from Swan of ATI Technologies [40], from Faroudja et al. of Faroudja Laboratories [14] and from Coombs et al. of Philips Electronics UK Limited [8].

Encoding for compression In MPEG encoding, flags are available that indicate that an image is a repeat of a previous image [22]. When a sequence is in 3:2 pull-down film mode, one film frame is shown three times. Since two of these fields produced of this film frame have identical interlace phases, one in five fields, or 20% percent reduction in the to be encoded data can be achieved by setting the field repeat flags during MPEG encoding [22]. Here film detectors are used to indicate the repeated fields in video streams. A good example of such a film detector is proposed by Del Corso [11]. Other proposed implementations for coding are Casavant et al. of Thomson [4], Martin of Thomson Consumer Electronics [29], Martin et al. of Thomson Consumer Electronics [30], Yonemitsu et al. of Sony Corporation [48], Lee of Electronics and Telecommunications Research Institute, Daejeon [24], Yagasaki et al of Sony Corporation [46], Hui of STMICRO-Electronics Asia Pacific PTE LTD [26] and Lim of LG Electronics [25].

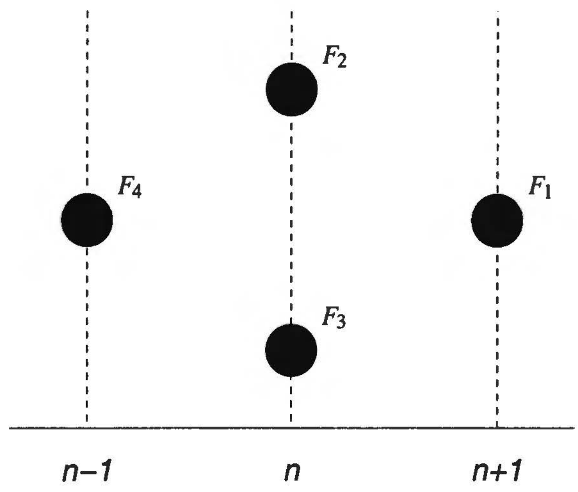


Figure 3.5: Kernel used by Correa and Schweer and Christopher and Correa.

The field of picture rate conversion or standards conversion is not included in this list. The methods applied in this field are film phase detectors and not full-film detectors, since they assume that the film mode of the incoming signal is known. An example of a film phase detector is proposed by Bock [3].

3.7 Reducing false detections due to static vertical detail

Difference detectors based on field memories also have a disadvantage compared to detectors using a frame memory. Detectors using a field memory have to deal with the interlacing of the video signal. Interlacing potentially introduces alias between vertical frequencies and temporal frequencies [2]. This alias causes high vertical high-frequency components to be detected as motion. This subsection is about schemes to reduce false detections motions due to vertical detail.

3.7.1 Four-pixel processing

Correa and Schweer of Deutsche Thomson-Brandt GmbH [9, 10] and Christopher and Correa [5, 6, 7], propose a strategy, four-pixel processing. In this strategy, a forward and backward comparison is used to determine if detected motion is due to vertical detail or due to pull-down motion. Pull-down motion is motion processed by the pull-down process. Pull-down motion has, in both 2:2 and 3:2 pull-down, the property that motion that is present between fields n and $n-1$, is not present between fields $n-1$ and $n-2$. Most film detectors that detect 2:2 pull-down try to detect this characteristic in the temporal patterns produced by the pull-down detector.

The method determines the presence of motion twice. Using the values shown in equation 3.13 (figure 3.5), once forward using the kernel F_1, F_2 and F_3 and once backward using the kernel F_2, F_3 and F_4.

Using this forward/backward motion detection, the presence of static vertical detail can be detected. In table 3.3 we see the processing of the motion signals.

A If both the motion detectors detect no motion, the decision is that there is no motion detected.

B If only the forward detector detects motion, the decision is that pull-down motion is detected.

C If only the backward detector detects motion, the decision is that pull-down motion is detected. The phase of the pull-down motion is reversed.

D If both the detectors detect motion, the decision is that the detected motion is either caused by vertical detail or motion not processed by the pull-down process. In that case, the film detector ignores that motion.

	forward	backward	decision	grade
A	no-motion	no-motion	static area	0
B	motion	no-motion	pull-down motion	-1
C	no-motion	motion	pull-down motion	1
D	motion	motion	static vertical detail/normal motion	0

Table 3.3: Decision logic for forward/backward difference detector.

An additional advantage is that noise in this system will randomly cause the forward and the backward detector to detect motion. By grading a B detection with -1 and a C detection with $+1$, the effect of noise will, if a large number of values are tested, cancel itself out, because half of the time case B occurs and half of the time causes case C occurs. In contrast with zero vector matchers, no threshold or coring step has to be implemented to reduce the effect of noise.

3.7.2 Field motion masking by frame motion

Another method to reduce the effect of static vertical detail is proposed by Swartz of Faroudja Laboratories [41]. This scheme employs a combination of a frame difference based motion detector and a field difference motion detector. First, motion is detected in the video signal using the frame memory based motion detector. This detector does not suffer from vertical detail, since it does not have to cope with the interlacing of the incoming video signal. Areas where frame motion is detected are expanded in both spatial and temporal directions. The field based motion detector only detects motion in those areas, producing the motion signal that the temporal pattern detector uses to detect the mode of the incoming video signal. The frame based film detector filters out the areas where static vertical detail occurs.

3.7.3 Discussion

The detector proposed by Correa and Schweer [9, 10] only detects motion that has pull-down properties, it is blind for any other kind of motion. This can cause the detection of 2:2 pull-down, even in sequences with large amounts of motion in video mode. To counter this effect Correa and Schweer propose to detect pull-down patterns over 64 fields. This results in a slow response to mode changes and still does not completely prevent false mode detection caused by temporal disturbances.

3.8 Zoning

By employing zoning, the film detector processes only part of the image. By sub-dividing the image into several zones, a set of film detectors can run in parallel, each processing their own zone. To make sure that all motion is detected, the whole image has to be covered by a set of windows.

The output of these film detectors can then be processed to determine the film mode for the entire image. Smaller windows result in an increased sensitivity for composite video elements, e.g. a scrolling text announcing a special news bulletin during a movie. These video elements do not contain pull-down and will cause annoying artifacts if they are processed in film mode. Because the amount of motion of the composite video elements is usually much smaller compared the motion in the rest of the image, it is probable that these elements will not be detected by the film-detector.

Correa and Schweer of Deutsche Thomson-Brandt GmbH [10, 9] propose such a scheme. Here, the image is divided into four horizontal sections. In each section the mode of the video signal is determined. If all four sections agree that the mode in their zone is film, the overall mode of the video signal is determined as that film mode. Using this method, a composite video element in one of the zones will be detected more easily and make the film detector to avoid incorrect processing of the video signal, e.g. introducing artifacts by applying film mode processing to sequences in video mode.

Another application of windowing is the improvement of robustness of detection. Martin of Thomson Consumer Electronics [29, 30] proposes to divide the screen into horizontal strips. Each strip consists of eight horizontal lines. In each strip, the mode of the video signal is determined. If three strips detect film mode, the output of the whole image is determined as film mode. In this application, zoning is used to increase the sensitivity. In the case of hybrid images, the film detector will sooner detect film mode.

Another application of windowing is proposed by Hui of STMICRO-Electronics Asia Pacific PTE LTD [26]. Hui proposes to divide the fields into a raster of zones. The mode detection for the entire image is film if a sufficient number of zones detect film mode. This design is proposed to 'increase the flexibility and detection accuracy'.

The proposals by Martin and by Hui use zoning to increase the robustness of the detection, while Correa and Schweer use zoning to increase sensitivity for the detection of hybrid images.

3.9 Conclusions

In this chapter, an overview of the field of film detection is given. Although it's not complete, it does give us insight into the problems that are present in this field.

Of the two pull-down processes that are known, 3:2 pull-down seems (almost) always to be detected using an absolute difference measurement over a frame period. This eliminates the need for an embedded de-interlacing mechanism in the difference detector.

For 2:2 pull-down, a de-interlacing mechanism is required. These de-interlacing mechanisms can introduce serious artifacts. When using luminance values containing these artifacts in the difference measurement, they cause large differences. These differences affect the difference value δ_{dei}. These differences are incorrectly interpreted as motion. The challenge is to devise a difference measurement that can detect motion, without suffering from artifacts introduced during de-interlacing.

Besides the problem of *false motion detection*, described in the previous paragraph, the problems of *video composites* and *video edits* are two mayor sources of problems.

With video composites, the video signal contains multiple regions. These sequences are created by composing (part of) other sequences. These other sequences can originate from different sources. This results in a *mixed film mode* sequence, where some parts are in video mode and some parts are in film mode. The detected film mode depends on the content of the sequence. Since both sequences are present, the chance of incorrect processing by the video chain, and thus the introduction of serious artifacts, is great.

With video edits, the converted film is post processed. In this post processing, the video signal is re-edited into a new group of sequences. This editing, that is done on a video signal, can result a break in the film pattern. The sequence that is put after the first sequence does not have to progress the film phase. A mismatch in phases causes a *break* in the repeat pattern, causing the temporal pattern detector to loose it's lock on the film mode. This will result in a new *run-in period*, where the film detector has to re-detect the film mode.

Two major application fields have been observed:

Frame rate conversion In this application, film detection is used to implement perfect de-interlacing by field insertion. It is important not to process sequences in video mode as sequences in film mode, as the field insertion de-interlacing technique can introduce serious artifacts. This demands a high quality of the film detector. Additionally, some frame-rate converters remove motion judder artifact, introduced during the telecine conversion process.

Compression During the encoding stage of the compression process, repeats of images are indicated as such. These repeats do not have to be encoded, as the data is already available in at decompression. This reduces the amount of data that has to be stored or transmitted. In this application, only 3:2 pull-down is detected, since only this type of pull-down has repeats of fields with the same interlace phase. A missed detection of the film mode does not directly lead to the introduction of artifacts, but to an increase of the amount of data that has to be encoded. This makes the quality constraint on the film detector less strict than in the field of scan-rate conversion.

Chapter 4

Schutten-Riemens Film Detector

In this report, we focus on an economic implementation of a film detector for the Tri-media video processor architecture [37, 38]. Within Philips Research, there is special interest in gaining knowledge of video processing on such programmable platforms.

The current architecture is the Philips high-end de-interlacer for high-end television sets, the SAA4992 (Falcon) video processor [35]. The Falcon also contains a mechanism for detecting the source of the video signal. This film detector is based on the analysis of the estimated motion vectors [19]. The motion estimator produces these motion vectors. This introduces a chicken and egg-problem. The motion estimator depends for its functioning on the film mode, that is generated by the film-detector. The film detector, in turn, uses the output of the motion estimator to determine the film mode. This counter dependency of two non-linear systems can cause the system to reach a deadlock situation in the film detector. In order to resolve such a deadlock situation, an egg-slice detector is incorporated into the design. This egg-slice detector is a crude jagged edge detector. It is designed to only resolve the deadlocked situation. The performance of this detector is too poor to be used as a stand-alone film-detector.

The Condor video scan rate conversion system [39] was designed as a successor for the Falcon video processor. To reduce complexity and increase the testability of the system, Condor architecture should resolve the counter dependency of the motion estimator and the film detector in Falcon. The solution came in the form of a stand-alone film detector. Condor's film detector, the Schutten-Riemens Film Detector (SRFD) is designed be developed and tested without the need to develop any other part of the video chain. This has the advantage that it can be transferred into other video processing architectures as it is designed and implemented as a separate component, in contrast with the Falcon film-detector.

The SRFD will be implemented in the Viper-II video processing system, which is based on the Nexperia platform and incorporates a Tri-Media CPU. This CPU will handle several tasks. One of these tasks is film detection. The SRFD analyzes the incoming video signal $F[n]$ (figure 4.1) and generates the film mode (M)and film phase (ϕ) signals. These signals are used as control signals in of the scan-rate converter. This

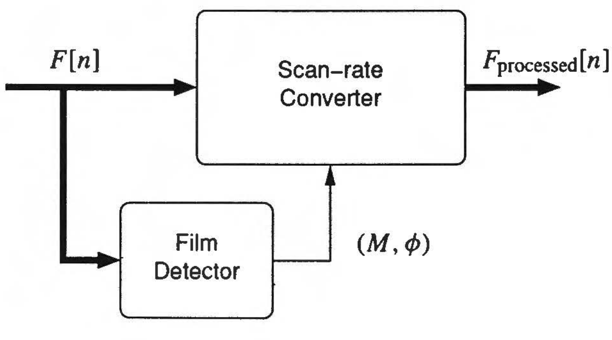


Figure 4.1: Placement of the film detector in the video processing chain.

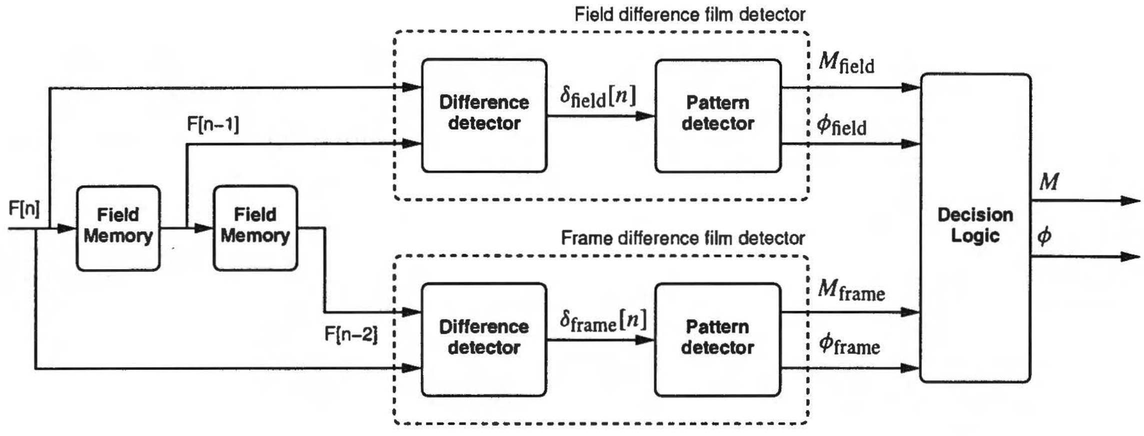


Figure 4.2: Structural over view of the SRFD.

scan-rate converter processes the incoming video signal $F[n]$. Based on the film mode signals (M and ϕ) it produces the outgoing video signal $F_{\text{processed}}[n]$.

The SRFD is not specifically designed for the Tri-media CPU architecture. Its algorithms should be portable to other platforms. The SRFD is currently implemented on such an architecture. This implementation allows us to test the algorithms in real-time. This real-time processing significantly decreases the amount of time spent on finding for sequences that cause problems with film detection. The SRFD uses methods already discussed in the prior-art. It represents, according to us the current state of technology. Therefore, we will investigate and improve on this design.

In the next section, a structural overview of the SRFD is discussed. In section 4.2, we examine typical motion patterns, to illustrate the operation of the film detector. In section 4.3, we summarize the conclusions.

4.1 Structural overview

The structure of the SRFD, figure 4.2, contains the following components:

Field Memories These video memories act as field delays, to generate the delayed versions of the video signal $F[n]$. The signal $F[n]$ is delayed by one field period, creating $F[n-1]$ and is delayed two field periods (or a frame period), creating $F[n-2]$.

Field difference film detector This film detection path compares $F[n]$ and $F[n-1]$ to calculate a field difference $\delta_{\text{field}}[n]$. This field difference $\delta_{\text{field}}[n]$ is an estimate of the amount of motion in the sequence. The pattern detector generates a film mode M_{field} and a film phase ϕ_{field} signal for the field difference path, based on a history of the field differences $\delta_{\text{field}}[n]$.

Frame difference film detector This film detection path compares $F[n]$ and $F[n-2]$ to calculate a frame difference $\delta_{\text{frame}}[n]$. This frame difference $\delta_{\text{frame}}[n]$ is an estimate of the amount of motion in the sequence. The pattern detector generates a film mode M_{frame} and a film phase ϕ_{frame} signal for the frame difference path, based on a history of the frame differences $\delta_{\text{frame}}[n]$.

Decision Logic The film mode and film phase signals of the frame difference and field difference film detection paths are subjected to a set of rules, generating an overall film mode M and film phase ϕ signal.

The film detector has a 'Hybrid structure'. It combines a frame difference and a field difference based film detection path. Both film detectors generate a film mode and phase. Film detectors work with the assumption that, using the difference numbers generated by the difference detectors, an indication of the amount of motion in the sequence in a field can accurately be determined. Using this estimate of the

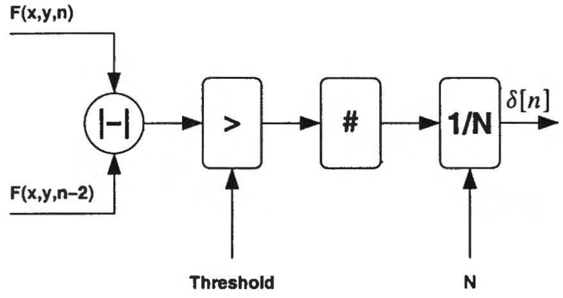


Figure 4.3: Structural overview of the frame difference detector.

amount of motion, a film mode and phase are determined in both film detection paths. Decision logic then generates a final film mode and phase based on the modes and phases from both the film detectors.

4.1.1 Hybrid Structure

A hybrid structure utilizes the advantages from film detection based on both field and frame difference patterns. While the field difference detector can detect both 3:2 pull-down and 2:2 pull-down film mode patterns. The reliability of this detection is lower than, that of a detection based on frame difference patterns. The frame difference detector, however, cannot detect 2:2 pull-down film mode. The hybrid solution takes the advantages of both the detectors. The final mode and phase are determined by decision logic. In this decision logic the following rules are implemented:

- If the frame pattern shows a 3:2 pull-down pattern, that output is used as the final result.
- If the frame pattern does not show 3:2 pull-down, the mode detected in the field difference pattern is used as the overall output.

4.1.2 Frame difference detector

The frame difference detector (figure 4.3) uses a zero vector matching on a frame difference scheme. The difference value $\delta[n]$ is calculated as:

$$\delta[n] = \frac{1}{N} \underset{\vec{x} \in A(n)}{\#} (\delta_{\text{frame}}(\vec{x}, n) > \text{Thr}) \tag{4.1}$$

with:

$$\delta_{\text{frame}}(\vec{x}, n) = |F(\vec{x}, n) - F(\vec{x}, n-2)| \tag{4.2}$$

where $A(n)$ is the set containing all the spatial positions that are checked for field number n, N is the number of elements in $A(n)$ and Thr is a threshold.

The detector takes the absolute difference of the luminance value $F(x, y, n)$ from the spatial position (x, y) in field n and $F(x, y, n-2)$, at the same spatial position in field $n-2$. A threshold is applied to this difference value to reduce number of differences generated by noise in the video sequence. The level of this threshold depends on the noise level of the sequence. The number of differences that exceeds the threshold is counted for a set of spatial positions in a field. This count is normalized by dividing it by the total number of compared differences N, producing a normalized difference number $\delta[n]$ for the field n.

The temporal pattern detector can, using a history of these normalized frame difference numbers $\delta[n]$, detect 3:2 pull-down frame difference patterns.

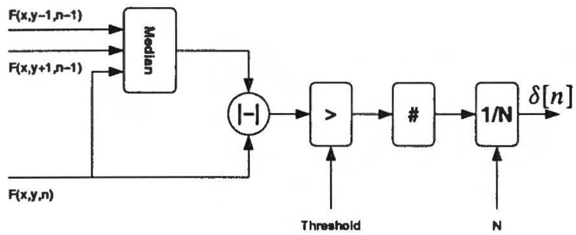


Figure 4.4: Structural overview of the field difference detector.

4.1.3 Field difference detector

The field difference detector (figure 4.4) uses a zero vector matching on a field difference using a de-interlacing scheme. The difference value $\delta[n]$ is calculated as:

$$\delta[n] = \frac{1}{N} \underset{\vec{x} \in A(n)}{\#} (\delta_{\text{dei}}(\vec{x}, n) > \text{Thr}) \tag{4.3}$$

with:

$$\delta_{\text{dei}}(\vec{x}, n) = |F(\vec{x}, n) - \text{med}(F(\vec{x}, n), F(x, y+1, n-1), F(x, y-1, n-1))| \tag{4.4}$$

where $A(n)$ is the set containing all the spatial positions that are checked for field number n, N is the number of elements in $A(n)$ and Thr is a threshold.

The detector takes the absolute difference of the luminance value $F(x, y, n)$ from the spatial position (x, y) in field n and a de-interlaced luminance value $F(x, y, n-1)$, at the same spatial position in field $n-1$. This de-interlaced value is calculated using a 3-taps median function on the value $F(x, y, n)$, $F(x, y+1, n-1)$ and $F(x, y-1, n-1)$. A threshold is applied to this value to reduce the number of detected differences caused by noise. The level of the threshold depends on the noise level of the sequence. The number of differences that exceeds the threshold is counted for a set of spatial positions in a field. This count is normalized by dividing it by the total number of compared differences N, producing a normalized difference number $\delta[n]$ for the field n.

Using a history of the normalized field difference numbers $\delta[n]$, the temporal pattern detector can detect 2:2 and 3:2 pull-down patterns.

4.1.4 Temporal Pattern detector

The temporal pattern detector detects pull-down patterns in a history of difference numbers generated by the difference detectors.

The pattern detector employs the strategy to stay in a certain mode until there is a reason to switch to another mode. This avoids a run-in period when a sequence in film-mode temporarily shows no motion. This is referred by Swartz [41] as 'the swinging pendulum problem'. The pattern detector analyzes a history of difference numbers to determine a film mode and phase.

4.1.5 Active window and sub-sampling

To reduce the cost of the implementation, a window and sub-sampling scheme is applied. The number of CPU-cycles spend on calculating the temporal difference patterns, is proportional to the number of checked differences. The number of checked differences is proportional with the area of the image that is checked.

To reduce this area, the active window, the SRFD disregards the areas of that do not belong to the active video signal. Additional to this, the size of active window is reduced by reduced by disregarding the parts in the over scan area of the image. Artifacts in the over scan area of the video signal are (generally) not seen by the viewer, because these parts are not visible on the display.

Applying an active window also eliminates differences caused by signals that are not part of the active video, e.g. teletext lines.

In order to reduce the number of CPU-cycles more, a sub-sampling grid is applied on the active window. We assume that, this sub-sampling covers the areas with motion and still results in enough samples to robustly determine the film mode.

4.1.6 Discussion

The SRFD is an implementation designed for robust film detection at a minimum cost. The application of the SRFD, being scan-rate conversion, adapts its de-interlacing scheme based on the detected film mode. This type of application asks for the detection of all progressively generated, i.e. non-interlaced content, e.g. computer generated animations.

The SRFD utilizes motion patterns to detect the film mode of a sequence. This implies that no robust detection can be made in the absence of motion in the video signal. This, however, does not pose a problem, since scan-rate conversion schemes are generally robust for sequences without motion. The film mode is not required for the conversion of static images. The problem lies in sequences with a small amount of motion. This, in combination with an amount of noise hinders the detection of the film mode. Noise added after the conversion process to video, is interpreted in the SRFD as motion in video mode. To reduce the effect of noise on the detection, a threshold is used. This threshold is, however, dependent on the amount of noise in the sequence. This requires an external device to measure the noise level. A robust noise independent solution would be preferred above such an implementation dependent on an external noise measurement device.

The interlacing of the video signal hinders the correct detection of 2:2 pull-down. The 2:2 pull-down process has the same temporal frequency as the interlacing process. This hinders the detection of 2:2 pull-down film mode. There is no duplicate information as used in 3:2 pull-down detection, where two fields with the same interlace phase can be compared directly. This absence of repeated fields forces the SRFD to deal with de-interlacing. De-interlacing of a video signal can introduce artifacts, when the film mode is unknown. These artifacts can cause large values in the field difference measurement. These differences are incorrectly interpreted as motion by the pattern detector.

Finally, the cost of implementation of the SRFD should be low. The implementation should run on a Tri-Media processor with a budget with no more than $10 \cdot 10^6$ processor clock cycles per second.

We expect improvements to be made mostly in determining the amount of motion in the sequence. Once this amount of motion has been determined accurately and reliably, the detection of a temporal pattern should be relatively simple.

For a detailed description of this detector, please refer to the Condor documentation [39][1].

4.2 Evaluation of Typical Behavior of the SRFD

To demonstrate the behavior of the film detector, we will investigate the motion patterns produced by the difference detectors. Using three typical sequences, we will show difference numbers by the current version of the SRFD. The film detector correctly detects the mode and phase of the sequences and we expect that they will show typical difference patterns. These three sequences are:

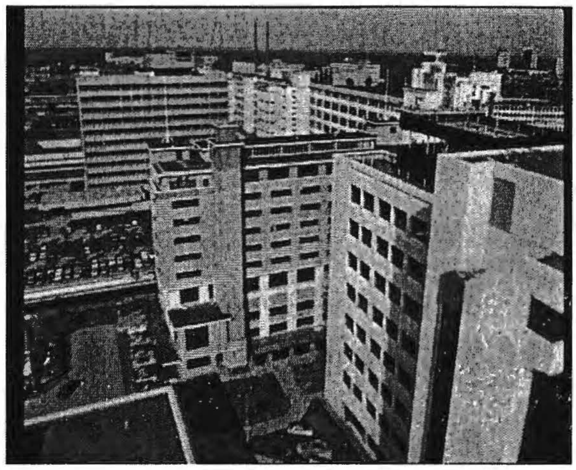
Renata This sequence (figure 4.5a) is in video mode. It originates from HD-MAC test recordings at RAI Ricerca Avanzata.

Flight This sequence (figure 4.5b) is in 2:2 pull-down film mode. It is a part of the promotional movie "Philips Expertise World Wide".

Usual Suspects This sequence (figure 4.5c) as in 3:2 pull-down film mode. This sequence originates from the 60 Hz version of the DVD of the motion picture "The Usual Suspects".

Next, we examine the `HL`-patterns, as discussed in section 3.4, in the field and frame difference numbers over a number of fields.

[1]The SRFD is referred in the Condor documentation as the Quality Film Detector (QFD).



a) b)

c)

Figure 4.5: Snapshots from the Renata sequence(a), Flight sequence(b) and Usual Suspects sequence (c).

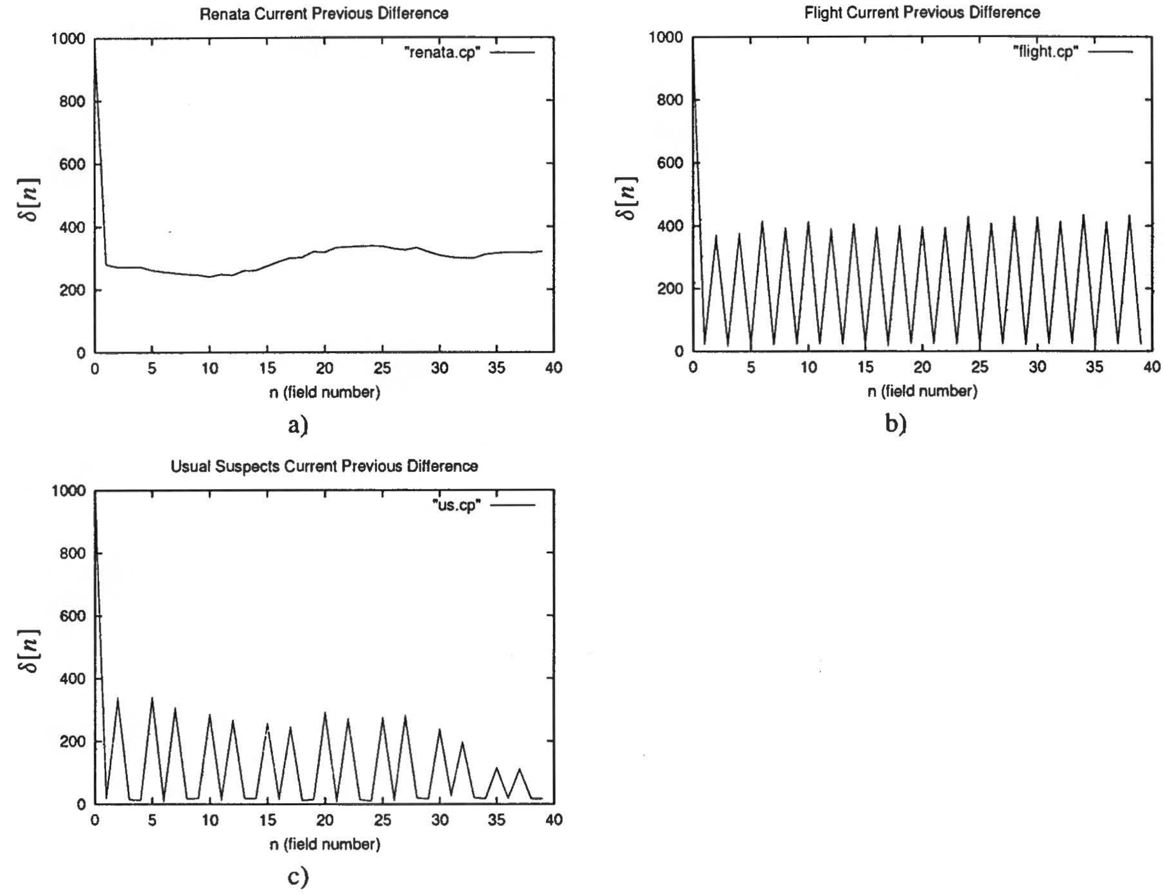


Figure 4.6: The difference numbers generated by the field difference detector. The horizontal axis is the field number; the vertical axis is the normalized amount of motion [0 ... 1000]. The difference numbers are generated from the Renata sequence (a), the Flight sequence (b) and the Usual Suspects sequence (c).

4.2.1 Field difference patterns

We expect that each of the three sequences will generate a typical field difference pattern as shown in the prior art. In the difference number we can observe:

- The Renata sequence shows a video field difference pattern. Observing the values presented in figure 4.6a), we can see that there is no alternating pattern of H and L values. The amount of detected difference changes gradually during the sequence. The pattern detector cannot detect a pattern and thus classifies the mode of this sequence as video mode
- The Flight sequence shows a HLHLHL pattern. This pattern is consistent with a 2:2 pull-down field difference pattern. This pattern can be seen in the values produced by the field difference detector, in figure 4.6b). The pattern detector classifies this sequence as 2:2 pull-down film mode.
- The Usual Suspects sequence shows a 3:2 pull-down field difference pattern. This pattern can be seen in the values produced by the field difference detector, in figure 4.6c). The pattern detector classifies this sequence is in 3:2 pull-down film mode.
- All three plots show high motion numbers at the first field difference. This is caused by the run-in of the algorithm. At the start, the image data is compared with a field filled with the value 0. This causes almost all examined differences to exceed the threshold.

Each of the observed field difference patterns corresponds with the expected patterns, mentioned in the prior art.

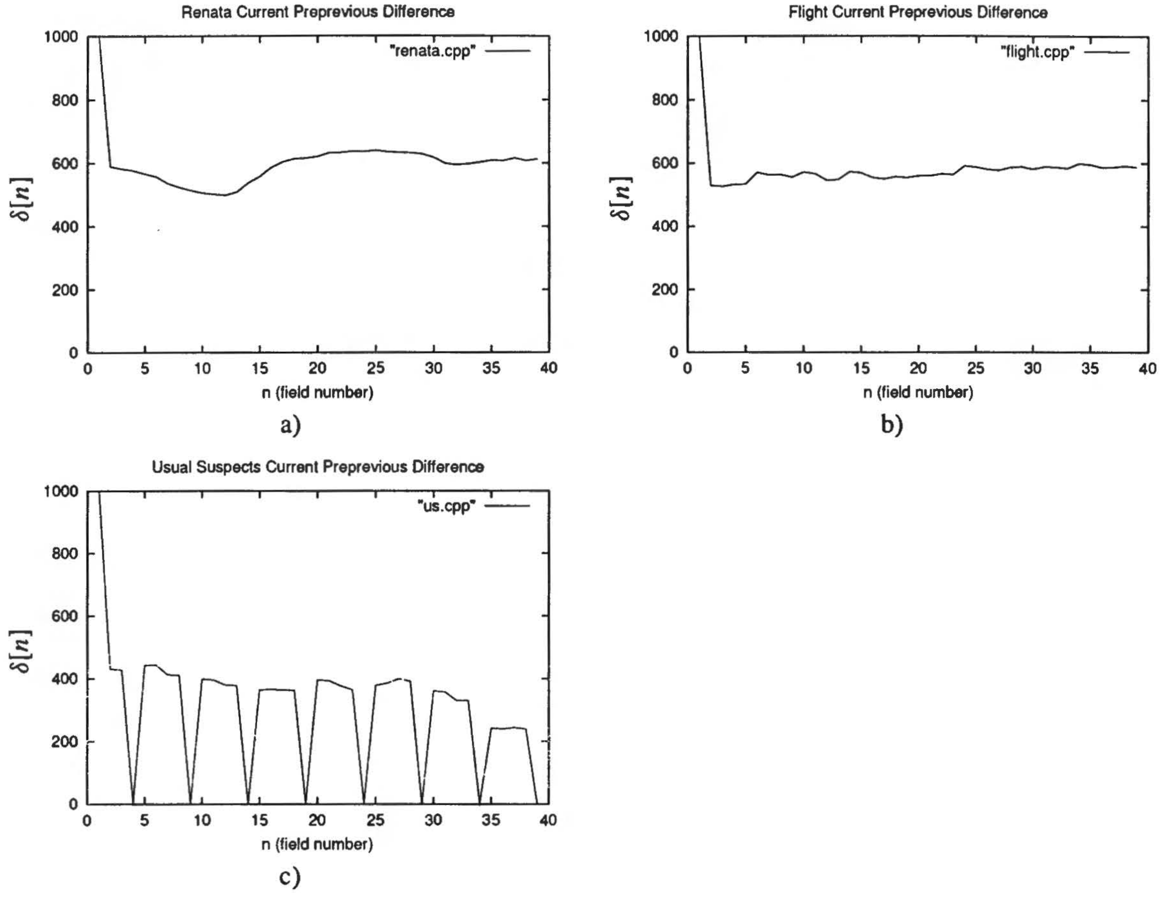


Figure 4.7: The difference numbers generated by the frame difference detector. The horizontal axis is the field number; the vertical axis is the normalized amount of motion [0 ... 1000]. The difference numbers are generated from the Renata sequence (a), the Flight sequence (b) and the Usual Suspects sequence (c).

4.2.2 Frame difference patterns

When examining the patterns generated by the frame difference detector, we expect that, for each of the three sequences, the field difference motion pattern corresponding to that type of sequence, can be observed.

- The Renata sequence shows a video field difference pattern. Observing the values presented in figure 4.7a), we can see that there is no alternating pattern of `H` and `L` values present. The pattern detector classifies this sequence as video mode.
- The Flight sequence shows no distinguishable `HL`-pattern, which is consistent with a 2:2 pull-down frame difference pattern. The pattern detector classifies this sequence as video mode.
- The Usual Suspects sequence shows a 3:2 pull-down field difference pattern. This pattern can be seen in the values produced by the frame difference detector, in figure 4.7c). The pattern shows a low value of every five fields. This is consistent with the expected frame difference pattern. The pattern detector classifies this sequence as 3:2 pull-down film mode.
- All three plots show high motion numbers at the first two field differences. This is caused by the run-in of the algorithm. At the start, the image data is compared with a field filled with the value 0. This causes almost all examined differences to exceed the threshold.

Each of the observed frame difference patterns corresponds with the expected patterns as documented in the prior art.

4.3 Conclusions

From the analysis of the typical sequences, we conclude that for these sequences, the SRFD is able to detect both the expected field difference patterns as the expected frame difference patterns for video mode, 2:2 pull-down film mode as 3:2 pull-down film mode. Using these field and frame patterns, the SRFD can successfully determine the correct mode and phase for these typical sequences.

Two additional observations were made. The amount of measured difference in the Renata sequence changes gradually. We explain this behavior by the property that the objects in the sequence (Renata herself and the background) show motion that is governed by inertia. This prevents sudden increases in the amount of motion.

The second observation was that *spikes* were observed in the difference numbers. Moving objects do not cause these sudden increases in motion. Some of these spikes occur at positions of *hard cut shot boundaries*. Other spikes occurred at the start of a sequence. Those are caused by the comparison of image data with field that are initialized with values set to zero. Comparing the data of the sequence with this data causes large differences to be detected, which show up as a spike.

Chapter 5

Problems with and improvements on the Schutten-Riemens Film Detector

This chapter contains proposals to improve the performance and extend the functionality of the Schutten-Riemens Film Detector (SRFD). These proposals are based on the observations made in sequences that are incorrectly classified by the film detector, even after a sufficient run-in period. In the next section we analyse several sequences where SRFD detects an incorrect film mode.

Based on those observations that are discussed in the next section, a new metric, the Arrow difference metric is introduced in section 5.2.1. The Arrow difference detector is designed to reduce the effect of static vertical detail. The alias from this static vertical detail causes the original difference metric of the SRFD to detect large differences. Those differences were incorrectly interpreted as motion. This can cause the film detector to detect the incorrect film mode.

In section 5.2.2, a hard-cut shot boundary detector based on the difference numbers is proposed. A shot is an uninterrupted segment of video frame sequence with static or continuous camera motion [32]. This detector can detect abrupt transitions between shots. We want to use this detector in the future to reduce the effect of video edits. This approach is based on the assumption that video edits create abrupt transitions between shots. Detecting these transitions will give us additional information about the presence of a video edit. In order to test this assumption, we require a hard-cut shot boundary detector.

The shot detector is part of a larger improvement. We do not include the shot detector in the evaluation, since this improvement on the SRFD is not fully developed. However, we do include it in this report, since the research on the hard-cut shot detector has lead to interesting results.

5.1 Problem Sequences Analysis

During testing of the SRFD, sequences have been identified where the SRFD detects an incorrect film mode. By examining these sequences, we determine the faulty behavior of the SRFD. These sequences are:

Fargo office This sequence (figure 5.1a) originates from the 50 Hz version of the DVD of the motion picture 'Fargo'. It is in 2:2 pull-down film mode. During the sequence the film detector detects parts in video mode.

Fargo ice plane This sequence (figure 5.1b) originates from the same DVD as Fargo Office. It is in 2:2 pull-down film mode. During the sequence the film detector detects parts in video mode, which is inconsistent with the expected mode of the sequence.

Ski text This sequence (figure 5.1c) is in 2:2 pull-down film mode. The SRFD decides on video.

RTL-z This sequence (figure 5.1d) is captured from live-broadcast from the Dutch TV channel RTL-5. It is in video mode. The sequence is sometimes detected as 2:2 pull-down. This seems to depend on the type of sub sampling of the examined luminance values used.

Parade This sequence (figure 5.1e) originates from a DV-video tape. The sequence is shot using a video camera. Parts of the sequence are detected as 2:2 pull-down film mode.

TMF This sequence (figure 5.1f) is captured from live-broadcast from the Dutch video clip channel TMF. The film detector detects 2:2 pull-down film mode and video mode during different parts of the sequence.

Each sequence has been analyzed. This analysis consists of an analysis of the motion patterns, in order to determine whether the incorrect classification is caused by the difference detector or by the pattern detector. If it is determined that the difference detector is the cause of an incorrect classification, the sequence is analyzed by 'tagging' the positions where the absolute luminance difference exceeds the threshold. Such a tag represents a contribution to the difference number $\delta[n]$ for this field. A custom built utility program 'filter_pfspd' tags these positions in a that field. A field of these tagged absolute differences forms a map that is used to examine the behavior of the difference detector. In the examples these tags are shown as white dots, superimposed on the original luminance values.

Figure 5.2 is an example of such a map of tagged differences. The fields originate from the Flight sequence. The sequence is in 2:2 pull-down film mode. Since this sequence is detected correctly, we assume that the maps created from this sequence will show typical behavior of the field difference detector.

The map in figure 5.2a shows the tagged differences between fields $n-1$ and n overlaid on the image in field n. The fields $n-1$ and n originate from the successive film frames. The map in figure 5.2b shows the tagged differences between fields n and $n+1$ overlaid over image $n+1$. The fields n and $n+1$ originate from the same film frame.

We observe that the map in figure 5.2a contains a large amount of tagged differences. These correspond with a large difference value $\delta[n]$. The map in figure 5.2b contains few tagged differences. These correspond with a small difference number $\delta[n]$.

By observing the tagged maps, we conclude:

- Two fields originating from successive film frames result in a high number of tagged differences and thus in a high or `H` field difference.
- Two fields originating from the same film frame have a small amount of tagged differences. This corresponds with a low or `L` field difference. We assume that this low non-zero value is caused by alias introduced in vertically detailed areas by the interlacing process.

The observations made in this section correspond with the expected values, described in section 3.4.

5.1.1 Problem Sequence I - Fargo office

The *Fargo Office* sequence is taken from the European 50 Hz version of the 1996 motion picture 'Fargo'. The SRFD should detect 2:2 pull-down film mode, for the length of the entire sequence. The sequence consists of three shots. The first and third shot are after a short run-in period detected correctly. The second shot, however is detected as video mode. Manual field-by-field comparisons show that the entire sequence is in 2:2 pull-down film mode.

Because we expect to detect 2:2 pull-down, we are only interested in the field difference motion numbers. The frame difference motion numbers are not used in 2:2 pull-down detection schemes. When we examine the field difference numbers in figure 5.3, we observe in the field difference numbers numbering:

0 - 1 A spike is present in the difference values at the start of the sequence.

1 - 44 The difference numbers show an alternating pattern of `H` and `L` values. The pattern detector correctly detects 2:2 pull-down film mode on this part, after a short run-in period.

44 - 45 There is a spike in the difference number history. This spike is at the position of a shot boundary.

45 - 198 The difference number pattern does not show an alternating `HL`-pattern. The pattern detector incorrectly detects video mode for this shot.

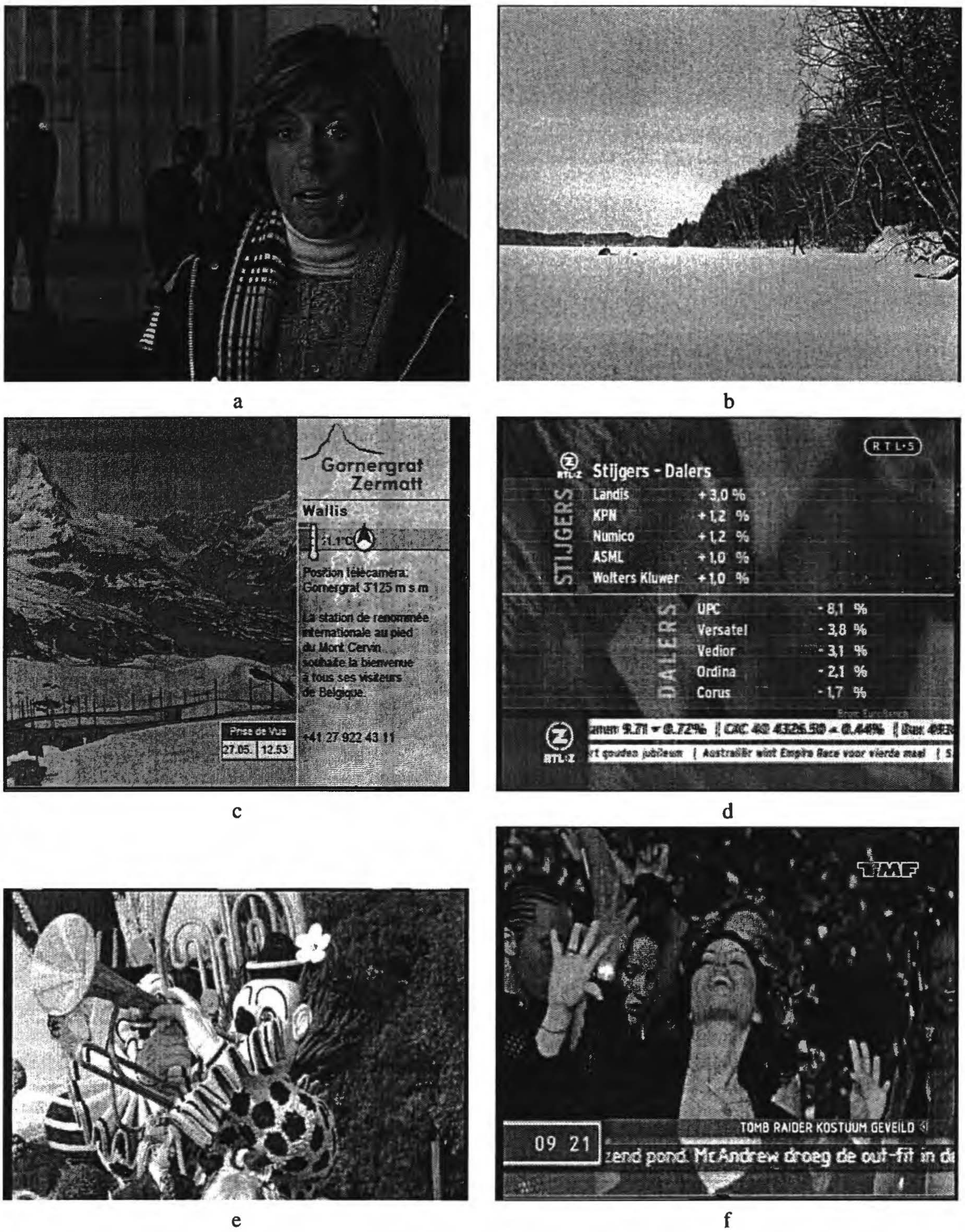
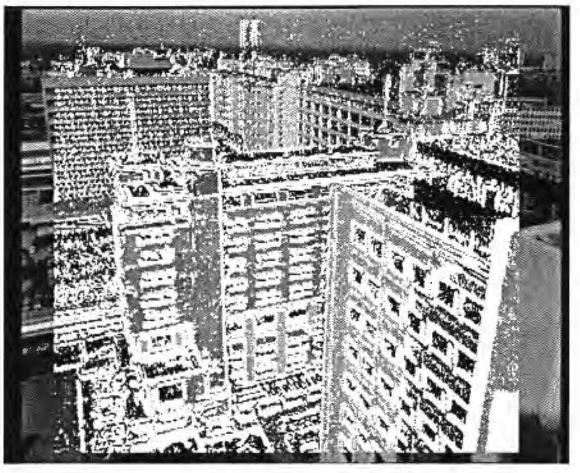
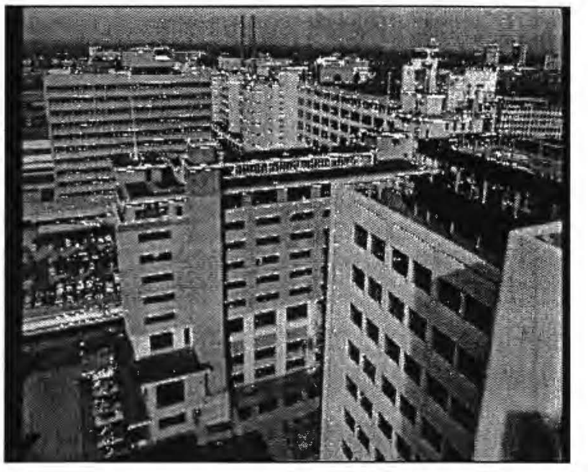


Figure 5.1: Snapshots from the sequences: Fargo Office (a), Fargo ice plane (b), Ski text (c), RTL-z (d), Parade (e), TMF (f).





a b

Figure 5.2: Two maps of tagged field differences superimposed on the current field, where the two fields originate from successive film frames (a) and from the same film frame (b).

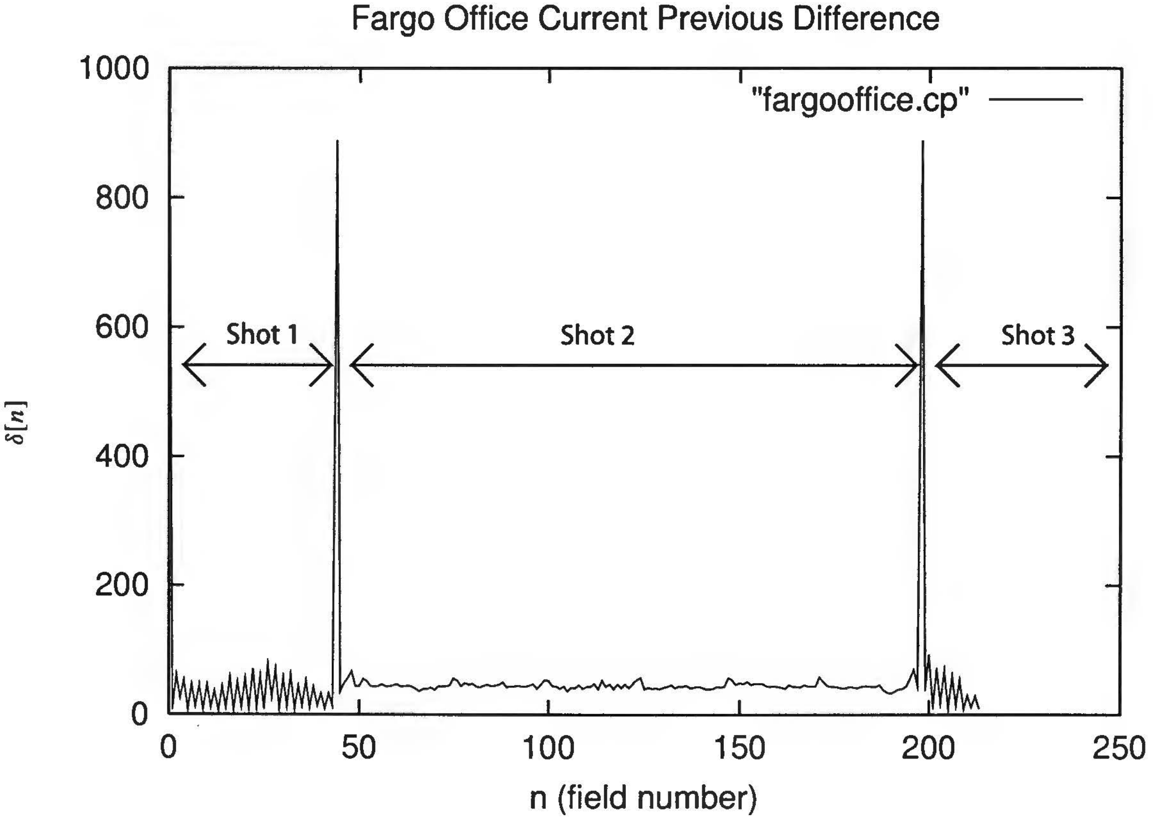


Figure 5.3: The field difference numbers generated by the field difference detector.

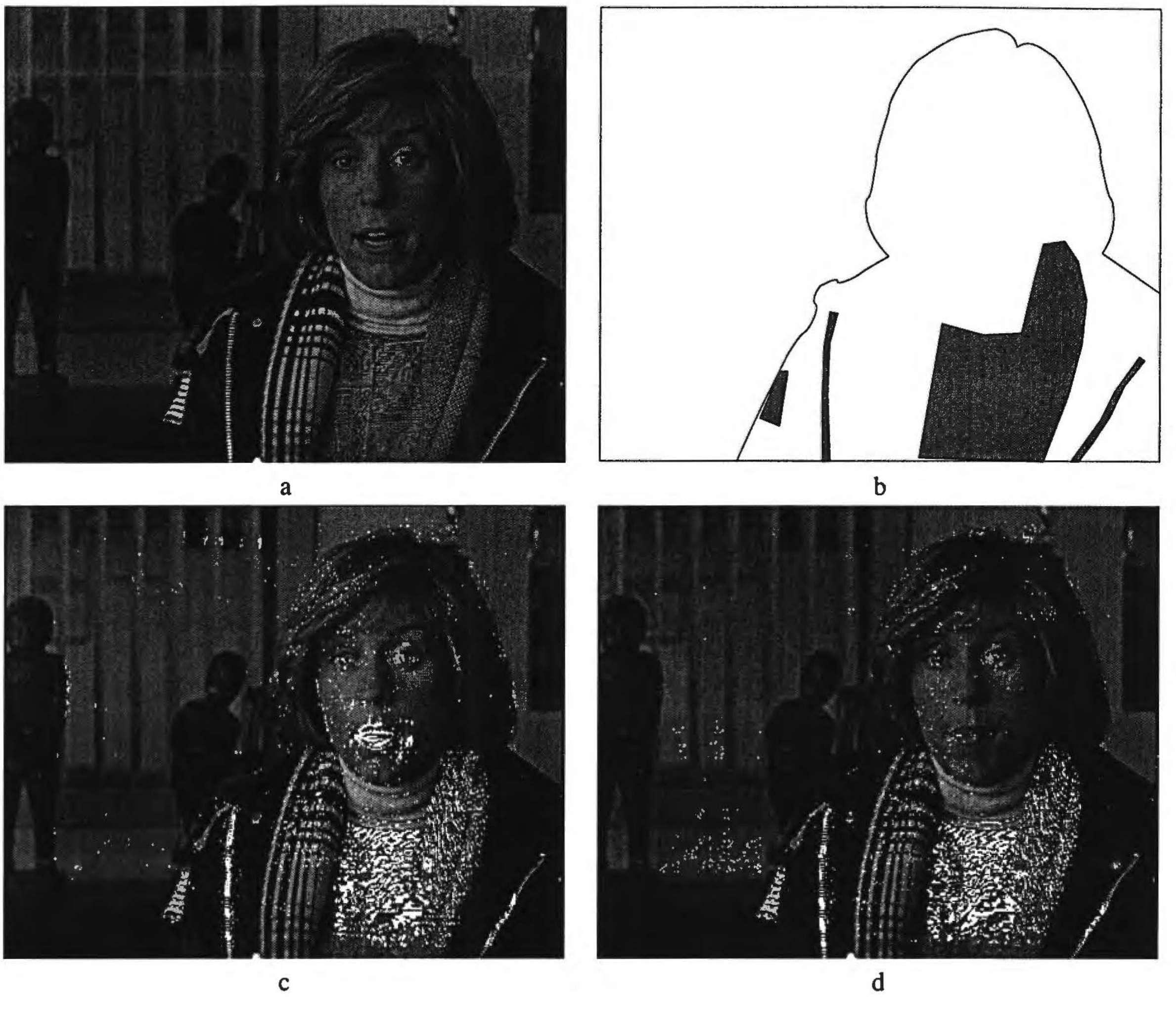


Figure 5.4: A screen capture of a frame from the second segment of the Fargo office sequence (a) and an overlay with the problem areas (b). Two maps of tagged differences, where the fields originate from different film frames (c) and from the same film frame (d).

198 - 199 Again there is a large spike in the difference number history. This is again at the position of a segment boundary.

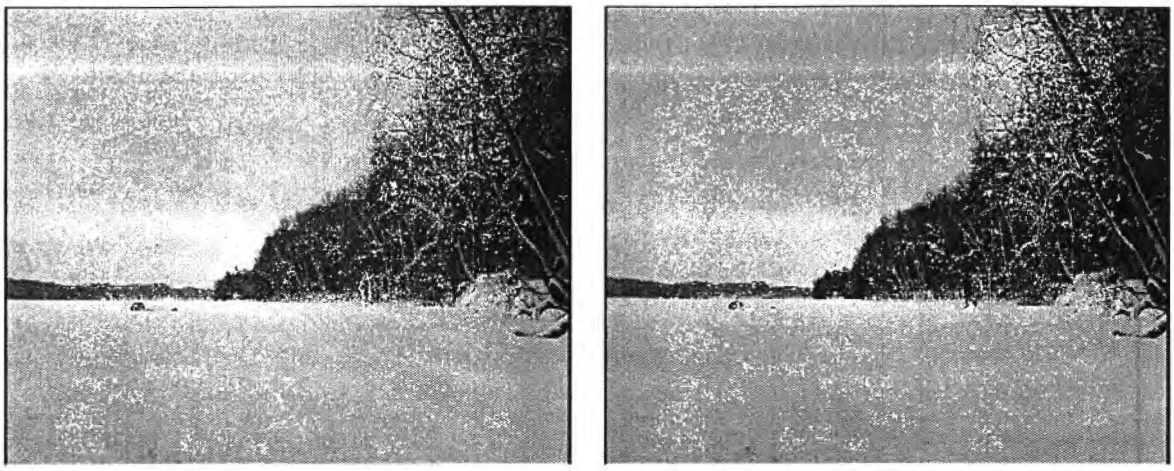
199 - 212 Here, a `HL` pattern starts. The pattern detector correctly detects 2:2 pull-down film mode, after a run-in period.

From this analysis, we draw the following conclusions:

- The field difference numbers in the field range 45 to 198 do show a pattern that is consistent with a field difference video mode pattern. The field difference detector is unable to produce field difference numbers that are consistent with 2:2 pull-down.
- Large field differences are observed at shot boundaries, compared to the difference numbers in the temporal neighborhood.

Since the field difference detector cannot generate a `HL`-pattern consistent with 2:2 pull-down film mode, the pattern detector can not detect the correct film mode. Here the field difference detector causes the incorrect detection.

In figure 5.4c and 5.4d, two maps of tagged field differences are shown. The map in figure 5.4c shows the differences between fields 118 and 119, each originating from successive film frames are tagged. The



a b

Figure 5.5: Two maps of tagged field differences superimposed on the current field, where the two fields originate from different film frames (a) and from the same film frame (b).

map in figure 5.4d shows the tagged differences between fields 119 and 120. We expect to see the a small amount of tags, as observed in the typical map in figure 5.2a. However, when we compare the amount of tagged differences in both maps, we see that they are about the same. This is consistent with the difference numbers generated from this sequence. In contrast with the typical map in figure 5.2b, the map in figure 5.4d contains a large amount of tagged pixels. This is *not* consistent with the expectations, where the amount of tagged pixels is low between two fields originating from the same film frame.

Closer examination of the areas where tagged differences occur in both the maps (the gray areas in figure 5.4b), reveals that these areas contain static vertical detail. We conclude that vertical detail in this sequence causes artifacts during de-interlacing. These artefacts can cause large differences in the field comparisons. Large differences between successive fields (forming a HHHH-pattern) are interpreted as characterizing a sequence in video mode.

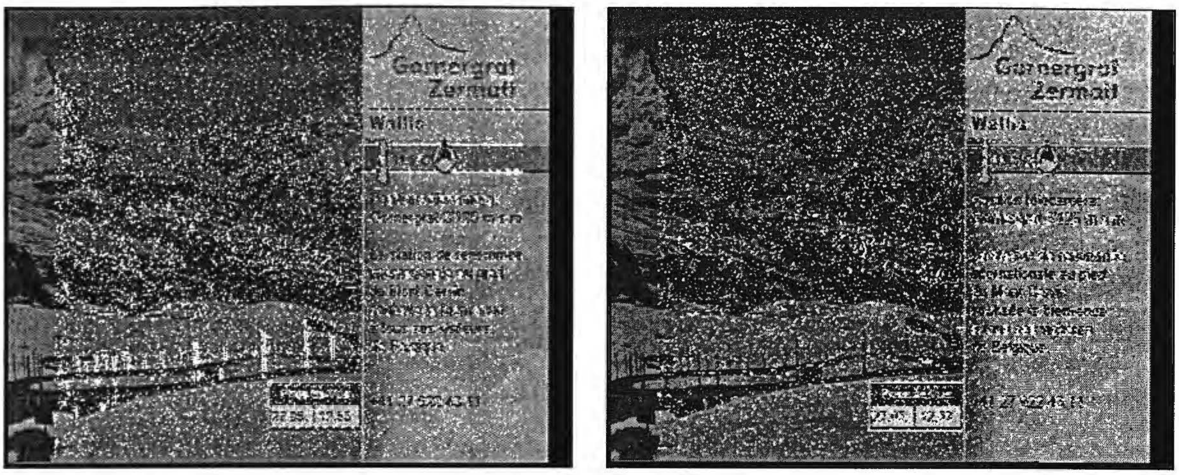
5.1.2 Problem Sequence II - Fargo ice plane

The *Fargo Iceplane* sequence originates from the same motion picture as the Fargo office sequence. Analog to the analysis method applied on the Fargo Office sequence, we have determined that the source of the incorrect detection of the film mode is the field difference detector. When we examine the mapped output of the film detector we see:

- that both the cases (figure 5.5a and figure 5.5b) are similar, producing similar maps of tagged differences.
- the moving policewoman in the center of the image is the only moving object in the sequences and produces the expected difference pattern expected a sequence in 2:2 pull-down.
- structured tagged differences in the trees on the right side of the image. These are probably caused by vertical detail.
- randomly distributed tags can be observed in parts of the image.

We suspect that these randomly distributed tags are caused by noise added after the transfer from film to video. Artifacts introduced during encoding can cause add to this differences. This can produce a sequence of numbers resembling a video mode difference pattern.

Noise and vertical detail produce a large amount of tagged differences. These are interpreted as motion by the pattern detector. Since the tagged differences caused by motion and vertical detail occur in each successive field difference, the pattern detector interprets these motion numbers as a HHHH pattern. This is consistent with video mode.



a b

Figure 5.6: Two maps of tagged field differences superimposed on the current field, where the two fields originate from different film frames (a) and from the same film frame (b).

5.1.3 Problem Sequence III - Ski text

The *Ski Text* sequence is taken from an information channel on skiing conditions. The sequence is a composite of a computer generated information panels and a panning view of the resort. This sequence is incorrectly detected as video mode, while the sequence contains areas containing 2:2 pull-down and static image content.

As with the previous examples, the difference numbers do not show a pattern consistent with 2:2 pull-down. This suggests that the difference numbers are caused by other sources than motion.

When we examine the tagged output of the difference detector (figure 5.6) we make the following observations:

- The image contains random distributed tags.
- Groups of tags can be identified around the lettering in the panels in both images.
- The image in figure 5.6a contains groups of tags around the poles in the mountains.

Again, we conclude that noise and vertical detail cause a large amount of detected difference between successive fields. These are interpreted as a `HHHH` pattern by the temporal pattern detector, producing video mode. The mountains and the poles show difference patterns consistent with 2:2 pull-down.

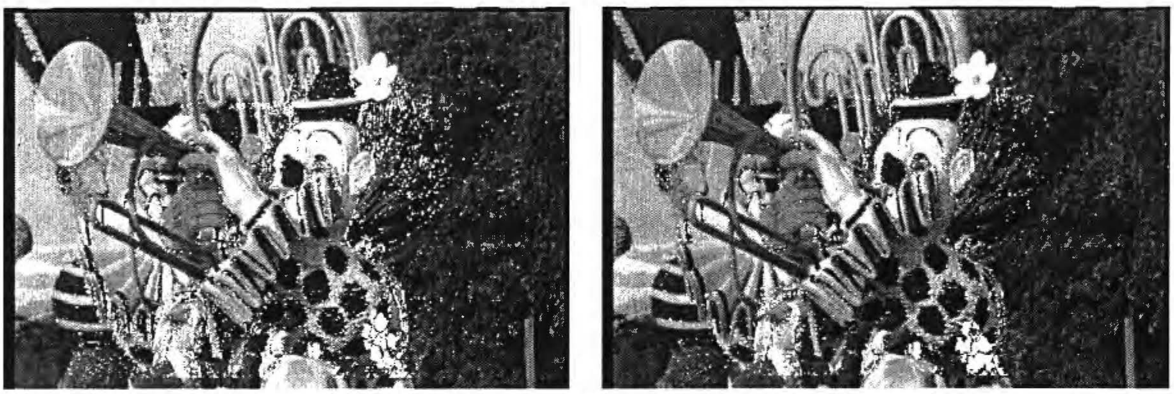
5.1.4 Problem Sequence IV - RTL-z

The *RTL-z* sequence is in video mode. It is correctly classified as video or incorrectly as film. This behavior seems to depend on the type of vertical sub-sampling used on the set of differences that are checked. When we examine the tagged differences in figures 5.6a and b we can observe:

- In each image there is a set of horizontal lines of tagged differences, though at different vertical positions. These lines are static in the sequence.
- Groups of tagged differences can be observed in the ticker at the bottom and around the text in the image.
- Randomly distributed tags can be observed in the entire image.

a b

Figure 5.7: Two maps of tagged field differences superimposed on the current field. Here map (a) shows the differences between fields n and $n + 1$ and map (b) shows the differences between fields $n + 1$ and $n + 2$.



a b

Figure 5.8: Two maps of tagged field differences, in a) between n and $n + 1$ and in b) between $n + 1$ and $n + 2$.

The incorrect detection of the film mode can be attributed to whether the sub-sampling covers the lines in just one of the two fields. The sub-sampling grid that covers these lines results in a significantly larger difference value than the result of a sub-sampling grid that does not these lines.

This causes that half the differences to be larger than. This alternating pattern of difference values is interpreted as a `HLHLHL`-pattern. Such a pattern is interpreted as 2:2 pull-down film mode.

The randomly distributed tags are probably caused by noise added during transmission.

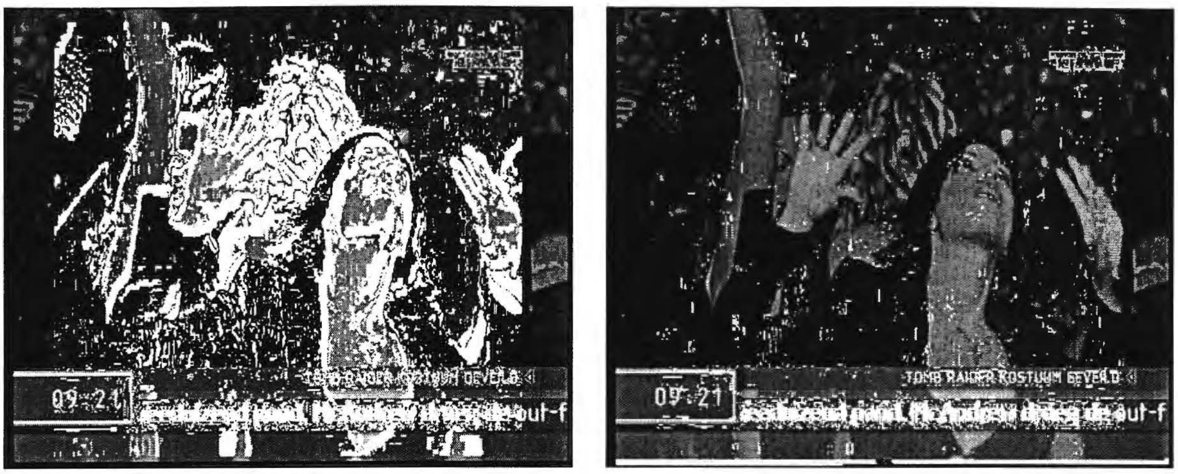
We conclude that vertical detail causes large luminance differences.

5.1.5 Problem Sequence V - Parade

The *Parade* sequence is recorded using a High-Definition Video camera. The sequence is converted to standard definition DV-video.

When we examine the tagged differences in figure 5.8a, we see that groups of tagged differences are visible around most moving objects in the sequence. However, when we examine the tagged differences in figure 5.8b, tagged differences can only be observed around some objects, but they are absent around other objects.

We suspect that the DV encoding causes parts of the image to be encoded as separate fields, while other parts are encoded as a frame. This reduces the amount of data required for this sequence, but introduces



a b

Figure 5.9: Two maps of tagged field differences, in a) between n and $n + 1$ and in b) between $n + 1$ and $n + 2$.

motion consistent with 2:2 pull-down patterns. This causes the sequence to be detected as a 2:2 pull-down film mode sequence, while the source is a video camera.

We conclude that there are objects in the sequence that produce both pull-down motion patterns as video motion patterns. The ratio between the amount of video motion and the amount of pull-down motion varies, causing unpredictable behavior of the film detector. The sequence cannot be correctly identified as film mode or video mode.

5.1.6 Problem Sequence VI - TMF

The *TMF* sequence was captured from a Dutch video clip channel 'The Music Factory'. The sequence is a video clip with an overlay containing a scrolling text. The SRFD switches between film and video mode during the sequence.

In the tagged maps, we can observe:

- The upper part of the screen shows tagged difference patterns consistent with 2:2 pull-down. We can see groups of tagged differences in figure 5.9a while these are absent in figure 5.9b.
- Tags appear in both maps around the clock and the 'TMF'-logo.
- Tags consistent with video mode appear around the scrolling text.

We suspect that the tags around the clock and the 'TMF'-logo are caused by static vertical detail in those areas.

This sequence contains parts that can be detected as 2:2 pull-down film mode and parts that are detected as video mode. During the TMF-sequence the film detector switches between 2:2 pull-down film mode and video mode. This sequence contains parts that exhibit the behavior of these modes. It cannot be correctly identified as film or video mode.

5.1.7 Conclusions

By examining the six problem sequences we observe that in those cases the erroneous detection is caused by an incorrect detection of motion. This incorrect motion detection is caused by:

Noise Noise causes differences to be tagged between fields.

Vertical Detail Vertical detail can cause artifacts during de-interlacing. These artifacts can result in large differences between fields. These differences can occur between every successive field difference.

This is interpreted as a HHHH-pattern by the temporal pattern detector. That results in the incorrect detection of video mode.

In the TMF and the Parade sequences, both video motion and pull-down motion has been observed. The SRFD cannot correctly classify these sequences. The SRFD assumes that the mode is a parameter that is valid for the entire image. This can cause problems when the film detector is used with an video-format converter, e.g. when a film mode is detected on such a sequence, field insertion de-interlacing is used, which can cause serious artifacts in areas that exhibit video motion.

To resolve this we propose to use a new mode: the *hybrid mode*. The hybrid mode signals the application that elements in the sequence exhibit both video and film motion patterns. This causes the application to use a fall-back strategy. In the application of video-format converters, this fall-back strategy would be to process the sequence as if it was a video mode sequence. This results in less annoying artifacts, which gives a higher perceived video quality.

Hybrid mode sequences are produces in various ways. A known way is by composite images (e.g. using overlays). Another source, discovered in this research, is encoding for compression (e.g. as used the DV-standard).

5.2 Improvements

5.2.1 The Arrow Difference Detector

In section 5.1, we pointed out that static vertical detail can cause difference numbers. These difference numbers can confuse the temporal pattern detector, causing it to detect the incorrect film mode.

In this section, we try to reduce the sensitivity of the film detector for the detection of differences caused by alias. This alias is introduced by static vertically detailed structures. We propose a difference metric that is insensitive to such alias. We suspect that the film detector fails to detect the correct film mode in the Fargo Office and RTL-z sequences due to this kind of alias.

For the film detector, we propose the following strategy: *Since we only need to take one decision per field on the film mode of the sequence, we can afford to ignore suspicious data.*

Using this strategy to we propose to utilize a static area detector that masks out areas that may contain static vertical detail.

For such a static area detector, we can utilize a frame difference measurement that is already available. Frame luminance differences do not suffer from interlacing since the two luminance values that are compared, always originate from fields with the same interlace phase. No alias is introduced in areas with static vertical detail. The frame difference, however, cannot detect 2:2 pull-down.

Using field difference measurement, we assume that: *If there is no signal from the frame difference, a field difference signal cannot be generated by motion.*

Any signal in the field difference, while motion on the frame difference is absent, must be caused by a different source than motion.

Using this assumption, we want to implement the *Arrow difference measurement.* The name Arrow originates from the shape of the kernel, where the wedge shaped kernel from the median function and the line formed by the kernel of the absolute frame difference form an arrow shape.

Analysis

In order to check the validity of our assumption, we investigate the spatial-temporal (f^y, f^t) spectrum. The spectrum of a static vertically detailed area lies completely on the f^y axis, since there are no temporal components in the area. When this sequence is interlaced, diagonal repeats appear [2] (figure 5.10). These repeats occur at half the temporal sampling frequency. The high frequent vertical components cause alias and appear as high frequent temporal components. These high frequent temporal components are interpreted as motion by the field difference detector.

We investigate the transition function of both the frame difference and the field difference measurement, in order to determine their behavior with regard to the alias caused by high frequent vertical components.

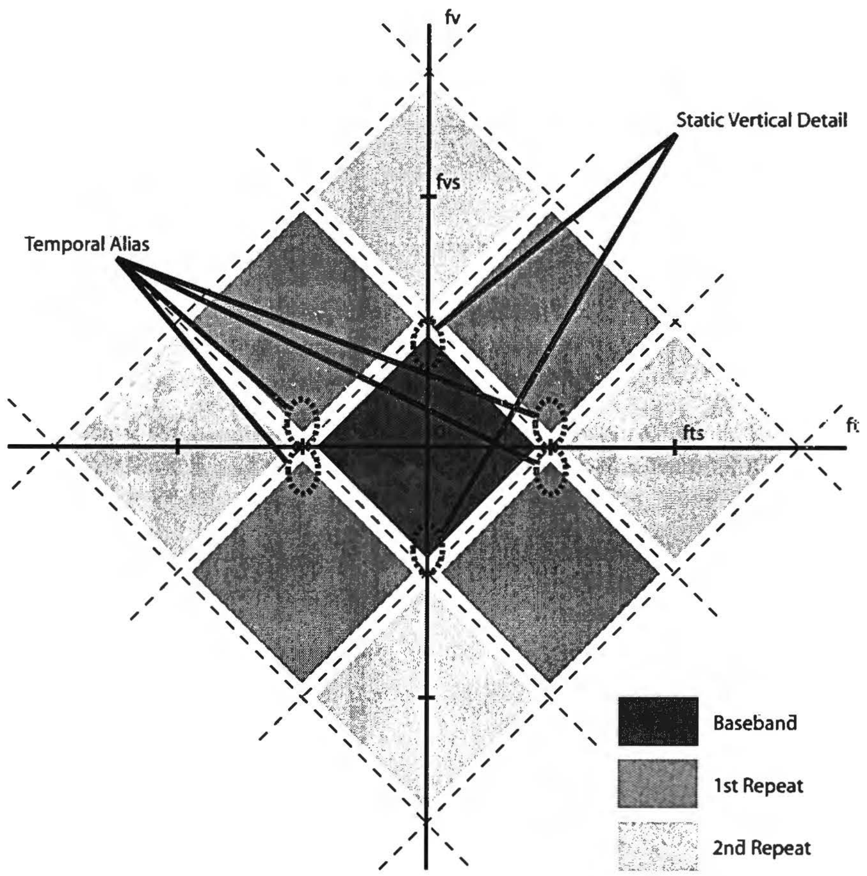


Figure 5.10: The spato-temporal spectrum of a quincunx interlaced signal, where fv is the vertical frequency, fvs the vertical sampling frequency, ft the temporal frequency and fts the temporal sampling frequency.

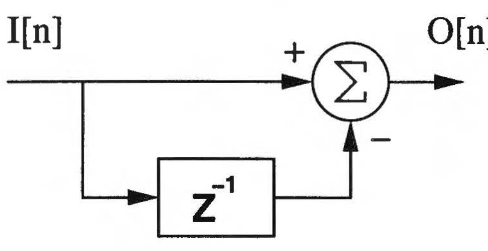


Figure 5.11: A schematic representation of a field difference measurement.

Field difference

The field difference measurement compares the current luminance value with a field period delayed luminance value (figure 5.11),

$$O[n] = I[n] - I[n-1] \tag{5.1}$$

where $I[n]$ is the input and $O[n]$ is the output of the difference function. The transfer function in the z-domain is

$$H_{\text{field}}(z) = \frac{O(z)}{I(z)} = 1 - z^{-1} = \frac{z-1}{z} \tag{5.2}$$

Where $O(z)$ is the z-transformed version of $O[n]$ and $I(z)$ is the z-transformed function of $I[n]$. The amplitude of the Fourier transform of this transfer function (figure 5.12) is:

$$\left|H_{\text{field}}(e^{j\theta})\right| = \frac{\left|e^{j\theta} - 1\right|}{\left|e^{j\theta}\right|} = |\cos(\theta) - 1| \tag{5.3}$$

where $z = e^{j\theta}$ and $\theta = 2\pi \frac{f^t}{f_s^t}$ with f^t the temporal frequency and f_s^t the temporal sampling frequency.

At half of the sampling frequency $f^t = \frac{1}{2} f_s^t$, $\theta = \pi$, the amplitude of the transfer function 5.3 at that frequency equals:

$$\left|H_{\text{field}}(e^{j\pi})\right| = |\cos(\pi) - 1| = |-1 - 1| = 2 \tag{5.4}$$

In the field difference measurement, the alias from static vertical detailed components causes large difference values.

Frame difference

Now, we consider a frame difference measurement (figure 5.13)

$$O[n] = I[n] - I[n-2] \tag{5.5}$$

where $I[n]$ is the input and $O[n]$ is the output of the frame difference function. The frame difference measurement compares a luminance value with a luminance value that that has been delayed twice. The transfer function in the z-domain of this frame difference is:

$$H_{\text{frame}}(z) = \frac{O(z)}{I(z)} = 1 - z^{-2} = \frac{z^2 - 1}{z^2} \tag{5.6}$$

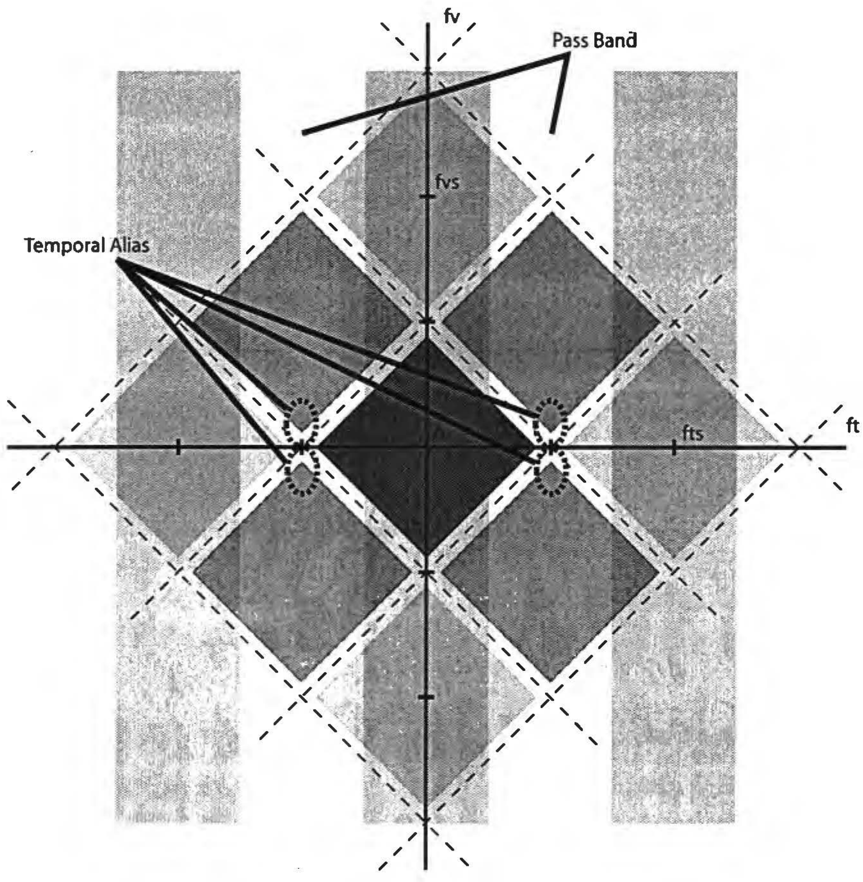


Figure 5.12: The temporal transfer of the field difference measurement, super-imposed on the spatio temporal spectrum, where fv is the vertical frequency, fvs the vertical sampling frequency, ft the temporal frequency and fts the temporal sampling frequency. This transfer does not suppress the temporal alias of the static vertical detail.

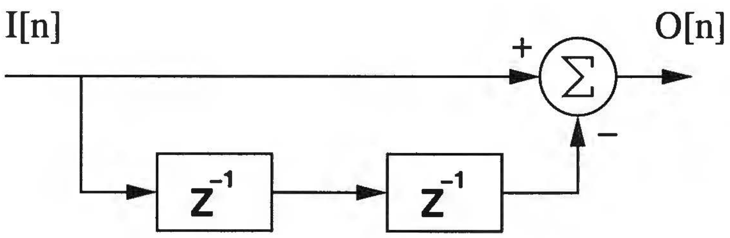


Figure 5.13: A schematic representation of a frame difference measurement.

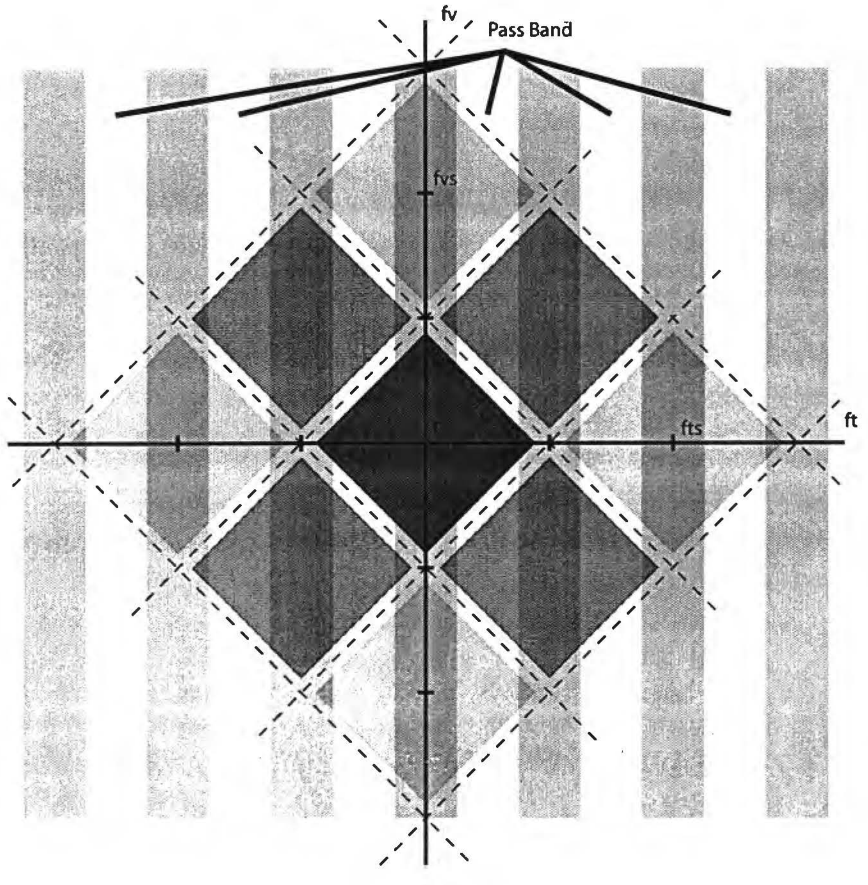


Figure 5.14: The Temporal transfer of the frame difference measurement, super-imposed on the spatio temporal spectrum, where fv is the vertical frequency, fvs the vertical sampling frequency, ft the temporal frequency and fts the temporal sampling frequency. This transfer suppresses the temporal alias of the static vertical detail.

Where $O(z)$ is the z-transformed version of $O[n]$ and $I(z)$ is the z-transformed version of $I[n]$. The amplitude of the Fourier transform of the transfer function (figure 5.14) is

$$\left|H_{\text{frame}}(e^{j\theta})\right| = \frac{|e^{j2\theta} - 1|}{|e^{j2\theta}|} = |\cos(2\theta) - 1| \tag{5.7}$$

where $z = e^{j\theta}$ and $\theta = 2\pi\frac{f^t}{f^t_s}$ with f^t the temporal frequency and f^t_s the temporal sampling frequency.

At half of the sampling frequency $f^t = \frac{1}{2}f^t_s$ or $\theta = \pi$, the amplitude of the transfer function 5.7 equals:

$$\left|H_{\text{frame}}(e^{j\pi})\right| = |\cos(2\pi) - 1| = |1 - 1| = 0 \tag{5.8}$$

In the frame difference measurement, the amplitude at half the temporal sampling frequency f^t_s is completely suppressed. The alias from static vertical detailed areas does not cause differences. However high frequent temporal components that are near the half the temporal sampling frequency f^t_s are also suppressed. The frame difference suppresses both high-frequent temporal components as the alias from high frequent vertical components in the spectrum.

Our strategy is to ignore differences that might not be caused by motion. We choose to ignore differences that are caused by high frequent temporal components, since we cannot distinguish them from

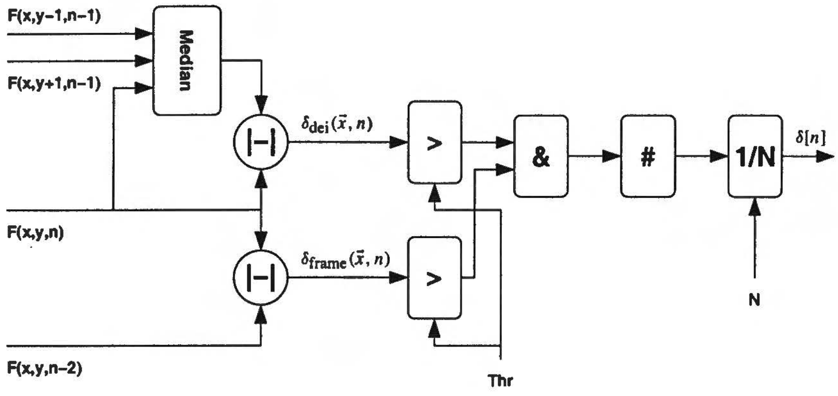


Figure 5.15: Structural overview of the arrow difference detector.

high frequent vertical components. We assume that, the contribution to the motion detection of these high frequent temporal components is small.

The Arrow difference measurement (figure 5.15) a logic AND port is used to combine the results of the thresholded field difference measurement and the thresholded frame difference measurement. This solution was chosen over a cascade of the two filters. The AND port allows for a more economical implementation, e.g. the $\delta_{\text{frame}}(\vec{x}, n)$ can be re-used for the detection of 3:2 pull-down.

We can conclude that a frame difference can successfully suppress differences caused by static vertical detail. Our strategy is to count only values for which we are certain that can show differences due to motion. We can afford to loose counts due to false masking. A strategy similar to this one is also implemented by Swartz [41] and is a common solution used in motion detectors. Swartz uses extends the aperture of the frame difference measurement in spatial and temporal direction. The field difference measurements at the spatio-temporal positions around the current position $(\vec{x}, n)$ are used to calculate the difference number $\delta[n]$. We chose not to do this, as it would significantly increase the cost of the implementation.

Proposed Solution

We propose an improved difference metric, the 'Arrow' dissimilarity measure:

$$\delta_{\text{arrow}}[n] = \frac{1}{N} \underset{\vec{x} \in A(n)}{\#} \big((\delta_{\text{dei}}(\vec{x}, n) > \text{Thr}) \wedge (\delta_{\text{frame}}(\vec{x}, n) > \text{Thr}) \big) \tag{5.9}$$

where

$$\delta_{\text{dei}}(\vec{x}, n) = \big| F(\vec{x}, n) - \text{med}\big(F(\vec{x}, n), F(x, y+1, n-1), F(x, y-1, n-1)\big) \big| \tag{5.10}$$

and

$$\delta_{\text{frame}}(\vec{x}, n) = |F(\vec{x}, n) - F(\vec{x}, n-2)| \tag{5.11}$$

The dissimilarity measure $\delta_{\text{arrow}}[n]$ counts the number of spatial positions in the set $A(n)$ for which $\delta_{dei}(\vec{x}, n)$ and $\delta(\vec{x}, n, n-2)$ exceed the threshold Thr and normalizes this by the total number of spatial positions N in $A(n)$.

Experiments

In order to test the proposed Arrow difference detector, we run the two problem sequences through the original difference metric and the new Arrow difference metric. We will compare the maps of tagged differences and the motion numbers for these sequences.

We compare the maps of the Fargo Office, RTL-z sequences produced by the arrow difference metric and the original metric. Furthermore, we will examine the maps produced by the metrics on a sequence of a static zone plate.

a b

Figure 5.16: The maps a and b are produced by the same fields as in figure 5.4c and d using the arrow difference measurement.

Fargo Office

The map in figure 5.4a and in figure 5.16a are created using the same fields as input. We can observe a significant reduction of the number of tagged differences in figure 5.16a, in the problem areas. There is still a large amount of tagged differences in areas that show motion.

The map in figure 5.4b and in figure 5.16b are created using the same fields as input. Here, we also can observe a significant reduction in the number of tagged differences.

The behavior of the difference metric is more consistent with the field difference pattern for 2:2 pull-down.

When we observe the motion numbers of the sequence produced by the Arrow difference metric (figure 5.17), we see that the amount of detected difference is significantly reduced. The behavior of the numbers (of almost no motion) is consistent with the motion in the shot.

Although the pattern detector cannot detect a film mode pattern, the differences are sufficiently low to stay in film mode. The whole sequence is now detected as containing no motion and the film detector remains in 2:2 pull-down film mode.

RTL-z

When we compare the output of the arrow difference metric with the original metric on the RTL-z sequence, we can see that the vertical lines of tagged differences in the original metric (figure 5.7a and b) are no longer present in the maps of tagged differences produced by the arrow metric (figure 5.18a and b). The motion in the scrolling text is still tagged. The pattern now always produces a pattern consistent with video mode. The pattern detector correctly detects video mode.

Zone plate

When we compare both metrics on a sequence containing a static zone plate, we can see that using the original metric, figure 5.19a, the alias from vertical detail causes a large number of differences to be tagged. These differences contribute to the field difference number. Using the Arrow metric, figure 5.19b, no tagged differences occur. This is consistent with the desired behavior of a difference metric, where areas without motion do not contribute to the difference numbers.

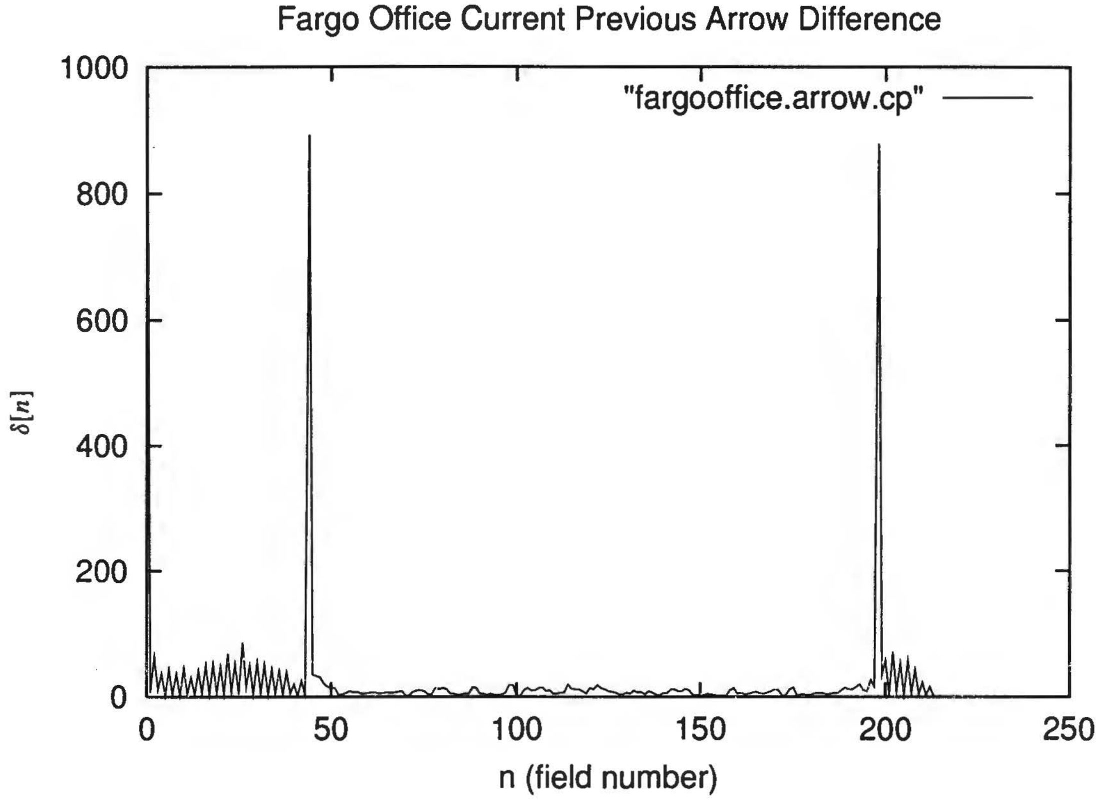
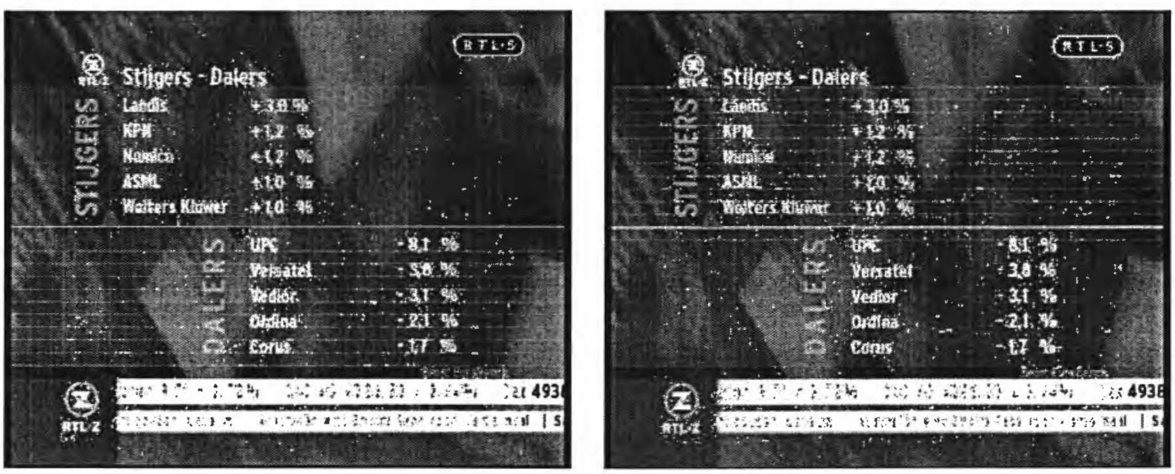


Figure 5.17: The field difference numbers generated by the arrow difference detector.



a b

Figure 5.18: The maps tagged differences a) and b) are produced by the same fields as in figure 5.7a and b using the arrow difference measurement.

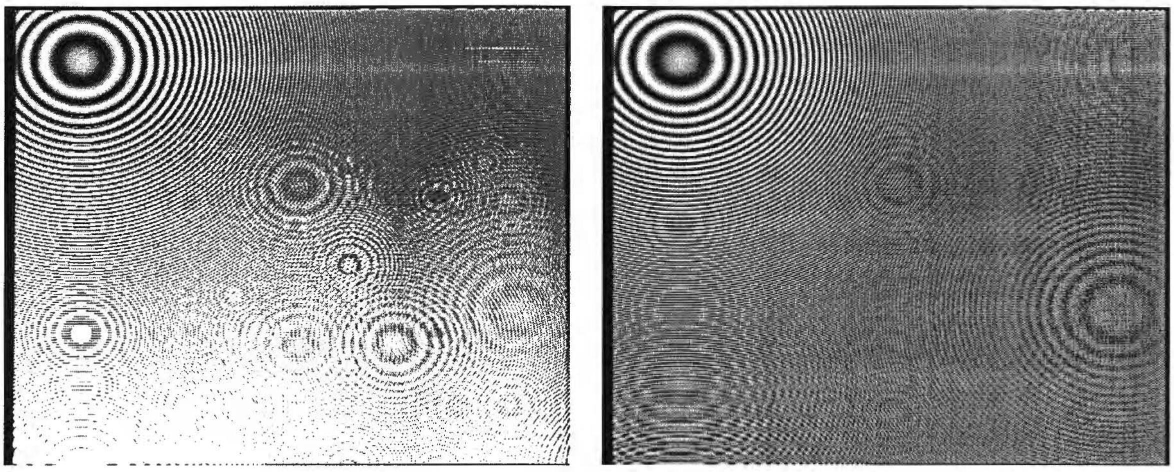

a b

Figure 5.19: The maps tagged differences of a static zone plate using the original metric a) and the arrow metric b).

Conclusions

From the measurements on the Fargo Office sequence, the RTL-z sequence and the zone plate sequence, we observe:

- The Arrow metric reduces the amount of differences detected due to static vertical detail in the Zone plate, RTL-z and Fargo Office sequences. The pattern detector interprets these differences as motion. Static vertical detail does not contain motion, and thus should not contribute to the difference number.
- The Arrow metric correctly can detect differences caused by motion in film mode in the Fargo office sequence and video mode in the RTL-z sequence.
- The Arrow metric detects no differences in the zone plate. The original difference metric causes a large number of differences to be tagged, while the whole sequence has no motion.

We conclude:

The Arrow detector is able to detect differences caused by motion in film mode and motion by video mode in the Fargo, RTL-z sequences. It is also able to ignore differences caused by static vertical detail.

5.2.2 Hard Cut Shot Boundary Detection

In order to reduce the number of false detections due to video edits, we propose to use a hard cut shot boundary detector. This approach is based on the assumption that video edits, an artifact where the difference pattern is disrupted, create hard-cut shot boundaries in the sequence. Knowledge about the presence of these shot boundaries, can give us an advantage in detecting these video edits. This could prevent a erroneous detection of the film mode after the video edit. In order to test this assumption we require a hard-cut shot boundary detector.

In this section, we propose such a method for hard cut shot boundary detection. This method uses the spikes observed in the field difference numbers (figure 5.4) at shot boundaries. We suspect that these spikes indicate the presence of a shot boundary. At a shot boundary, there is usually a sudden change in image content. Such a abrupt change in image content should cause large difference values. These large values should appear as spikes in the output of the difference detector.

For a shot, we use the definition of a shot by Ngo [32]. Ngo defines a shot as "an uninterrupted segment of video frame sequence with static or continuous camera motion, while a scene is a group of shots taken in the same site". Our detector is aimed at finding abrupt transitions between shots.

In the difference numbers observed in the 'Fargo office' sequence, spikes are observed at shot boundaries that are significantly larger than the numbers in the neighborhood of these shot boundaries. These large numbers are caused by the large difference in image content before and after the shot boundary. By detecting these large spikes, we assume that we can detect hard cut shot boundaries.

In the next section, we will discuss the field of shot boundary detection, then in subsection 5.2.2 we will propose methods for detecting shot boundaries using the numbers generated by the field difference detector. Next, these proposals are tested in a series of experiments. Finally, in subsection 5.2.2 we will draw conclusions based on the experiments.

Previous work on shot detection

In this subsection, we will give a short overview of the previous work in the field of shot detection.

Hanjalic [20] describes in what areas shot detection schemes are used. Shot detection is used in the areas of video-content analysis and content-based video browsing and retrieval. Here the shot is regarded as the fundamental element in the semantic structure of a large video sequence. For scene detection, shot boundaries are uses as a means to find the boundaries of scenes.

Lutong [44] describes the shot detector as a scheme consisting of a dissimilarity measure and a means to determine shot boundaries based on the output of that dissimilarity measure.

Hanjalic identifies two classes of shot boundaries:

Hard cuts These are abrupt transitions between two shots. The term cut originates from the motion picture industry, where an edit was made by cutting strips of film and sticking them back together in a different order. Each frame of the sequence belongs to a unique shot.

Gradual transitions This class consists of fades, wipes and other non-abrupt transitions. One shot evolves into a new shot over the course of several frames. During the transitions, the frames belong, to some extend, to both shots.

Lupatini et al [27] give three types of algorithms for shot boundary detection:

Histogram-based The histogram-based approach takes the benefit that histograms are insensitive to the amount of motion in the scene. However, the histogram-based approach fails to detect shot boundaries where the two shots have similar histograms.

Contour-based These are edge detection based shot boundary detectors. The contour-based approach make use of the ability in detecting gradual transitions.

Motion-based The motion-based approach is based on a motion based dissimilarity measure. These approaches are not favored because of the high sensitivity to motion. This sensitivity can cause large amounts of motion to be detected as a shot boundary.

Otsuji et al. [33] name types of sequences where problems occur for shot detectors. These types are Film, Slow motion and Animation. All three types use a type of pull-down scheme. The solution proposed by Otsuji uses a history of five fields to eliminate problems caused by pull-down. This history contains a value that is not affected by the pull-down process.

Additionally, Otsuji et al. show that luminance change based metrics are sensitive to motion, but insensitive to gradual changes, like fades. Histogram change based metrics are sensitive to gradual changes, but insensitive to transitions between two shots with similar histograms.

Proposed solution

From prior art we know that a shot detector usually comprises a dissimilarity measure and a means of detecting large dissimilarities. The SRFD has a dissimilarity measure in the form of the field difference numbers.

A global fixed threshold is a trade-off. A high threshold will have a small chance of false detections, but will have a higher probability to fail to detect a shot boundary. A low threshold will have a smaller chance of missing a shot boundary but a higher chance of false detections. To resolve this global trade off, we propose to use an adaptive threshold.

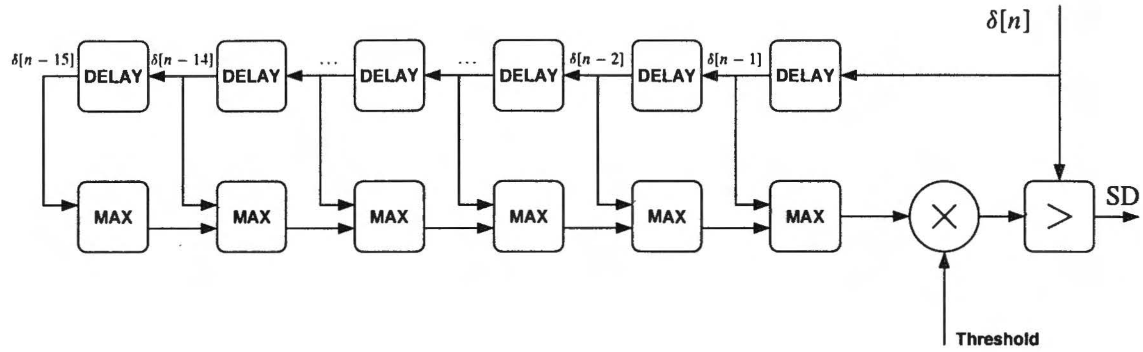


Figure 5.20: Structural overview of the hard cut shot boundary detector.

Such an adaptive threshold is based on the assumption that motion of natural objects is subject to inertia. Therefore, the amount of motion in a shot will change gradually. This effect can be observed in figure 4.6a, where the difference numbers change gradually. A sudden increase in the difference numbers can therefore not be caused by motion. Such an increase in the difference numbers must be generated by an event causing differences between the fields, other than motion. We assume that these large increases in field difference numbers are caused by hard cut shot boundaries. At such boundaries, the image content changes rapidly. This can cause large differences between fields, which in turn cause large difference numbers.

The detector detects a hard cut shot change when there is a significant increase in the amount of motion.

Sequences containing pull-down, such as animation, film and sequences in slow motion have an alternating pattern of high and low values. This pattern contains sudden increases and decreases of motion, caused by pull-down. In order to eliminate false shot boundary detections, the maximum of a history of difference values is used to compare with the new shot boundary.

This maximum reflects the maximum amount of motion. This should eliminate the effect of rapid changes caused by pull-down and other frame repetition schemes, when the maximum motion value is chosen from a history with sufficient length. The drawback of this scheme is, that a minimal detectable shot length is introduced. A previously detected shot automatically becomes the maximum value in the history. This maximum value does not represent the maximum amount of motion in the new shot. The spike of the previous shot change masks out the motion values. This causes a new sudden increase of the amount of motion to become undetectable.

We propose a scheme for detecting shot boundaries:

$$SD = \begin{cases} \text{true} & \delta[n] > \delta[m] + \text{thr} \forall m = n-1 \ldots n-15 \\ \text{false} & \end{cases} \tag{5.12}$$

where SD is a Boolean value indicating a shot boundary between the current frame n and $n-1$, $\delta[n]$ the measured difference value for the field n and thr the pre-set threshold value. The structure of such a detector is given in figure 5.20.

A shot boundary is detected if the current difference number is a higher than the sum of maximum value of a history of motion differences and a threshold. We propose to use the largest motion history currently used by the SRFD. This is a length of 16 field periods. The difference numbers are generated by the Arrow difference measurement.

Experiments

We perform on the proposed shot detector a set of experiments. These experiments are aimed at determining the behavior of the shot detector:

- with an increasing amount of added noise with a Gaussian distribution. This is to determine the robustness to noisy video sequences.

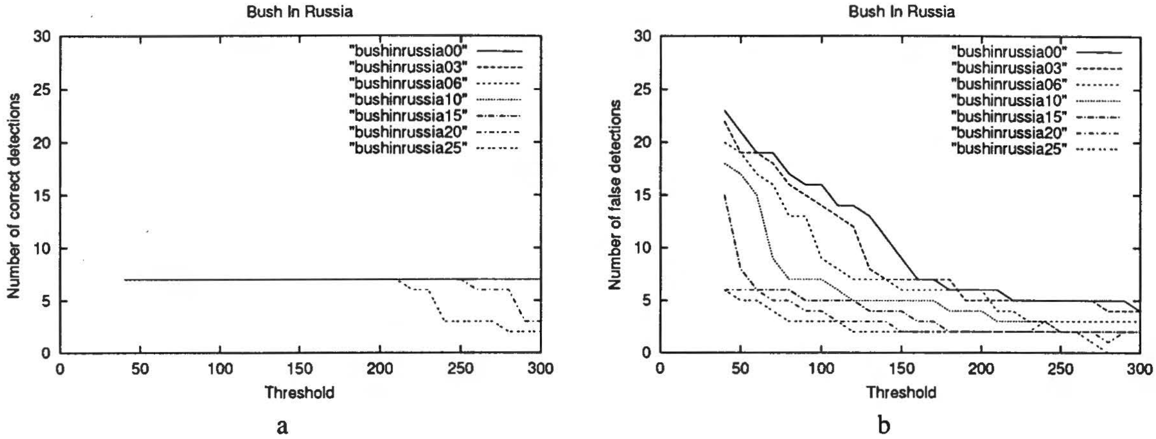


Figure 5.21: Bush in Russia: number of correct detections a) and false detections b) with increasing thresholds and increasing Gaussian noise levels.

- over a range of thresholds. This can be used to determine a threshold. The best threshold is dependent on the application of the shot detector.

Noise with a Gaussian distribution was added to the luminance values with standard deviations of 0, 3, 6, 10, 15, 20 and 25 is added to the 8-bit luminance values.

The output of the difference detector ranges from 0 (total similarity) to 1000 (total dissimilarity). The thresholds tested range from 40 to 300 with increments of 10.

The sequences used for this experiments are:

Bush in Russia This is a news broadcast. It contains flashes from photo cameras and is in video mode.

Gladiator This a sequence form the 'Gladiator' motion picture in 2:2 pull-down film mode. It contains a sequence with high amounts of motion.

Franklin This sequence is a cartoon. Cartoons contain irregular patterns of frame repetition.

Fargo This sequence contains part of a motion picture that has a moderate amount of motion. The sequence is in 2:2 pull-down film mode.

These sequences each have a length of 1800 fields. The shot boundaries have been recorded manually. These measurements show the number of:

Correct detections These are the hard cut shot boundaries that are present in the sequence and are correctly identified by the shot boundary detector.

False detections These are the hard cut shot boundaries that incorrectly identified by the shot boundary detector. They are not present in the sequence.

Observations

The results of the tests are presented in figure 5.21 for the sequence 'Bush in Russia', in figure 5.22 for the 'Gladiator Arena' sequence, in figure 5.23 for the sequence 'Franklin Cartoon' and in figure 5.24 for the 'Fargo' sequence.

We observe in figures:

5.21a 5.22a 5.23a 5.24a Each sequence shows a drop in the number of correctly detected shot boundaries with an increasing Gaussian noise level. Also we can observe that the number of correct detections drops with an increasing Gaussian noise level.

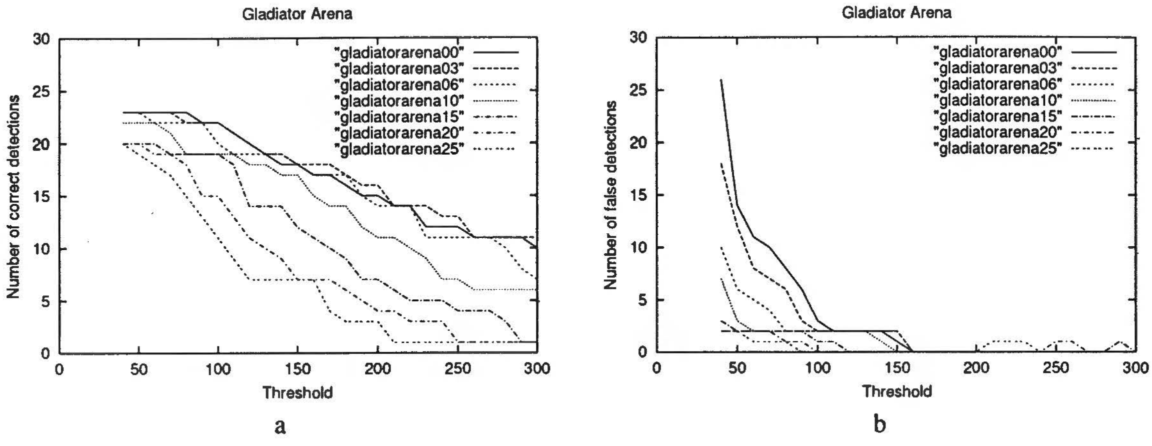


Figure 5.22: Gladiator arena: number of correct detections a) and false detections b) with increasing thresholds and increasing Gaussian noise levels.

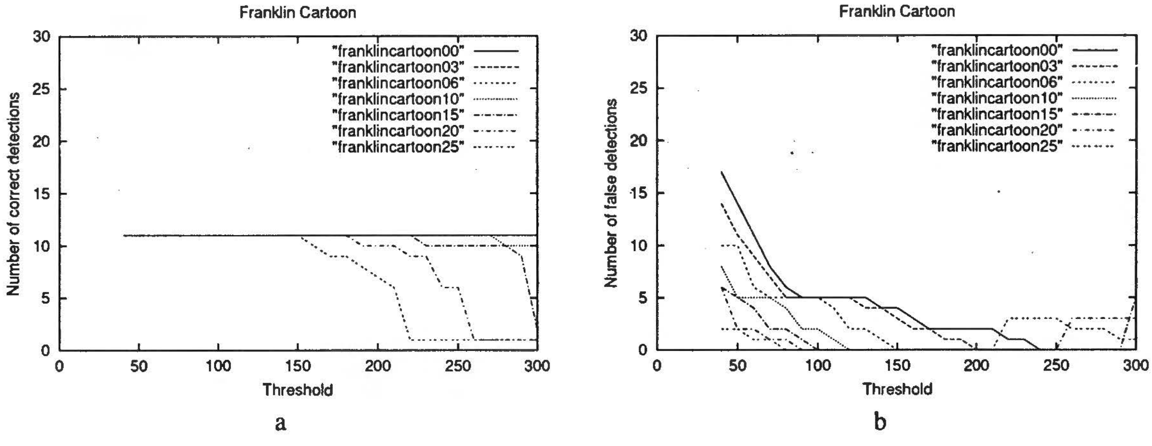


Figure 5.23: Franklin cartoon: number of correct detections a) and false detections b) with increasing thresholds and increasing Gaussian noise levels.

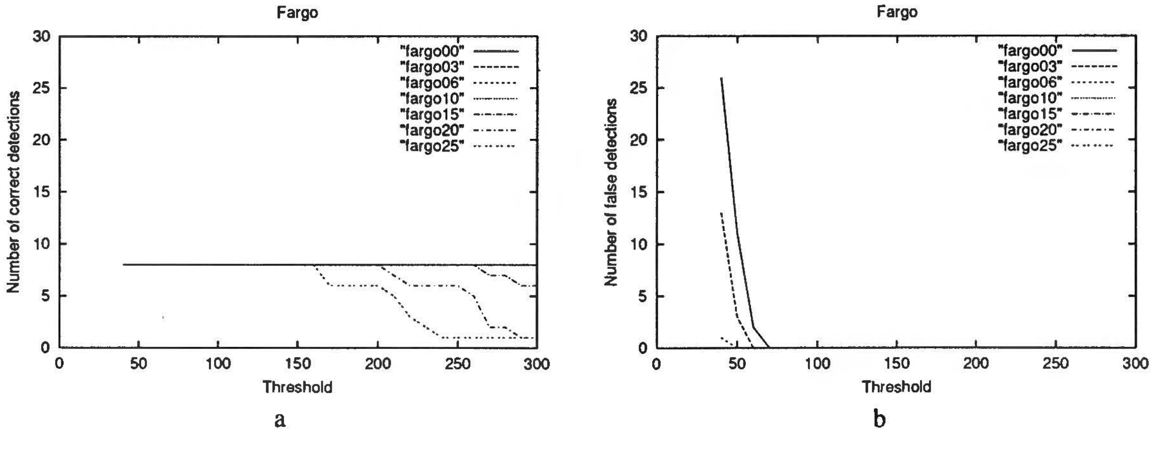


Figure 5.24: Fargo: number of correct detections b) and false detections d) with increasing thresholds and increasing Gaussian noise levels.

5.21b 5.22b 5.23b 5.24b We can observe that the number of false detections drops with an increasing threshold and that the number of false detections also drops with an increasing amount of added Gaussian noise.

In the plots of the number of correctly detected shots(figures 5.21a, 5.22a, 5.23a and 5.24a), we observe a decreasing number of correct detections with an increasing threshold. We suspect that this is caused a variation in the differences between the spike and the maximum difference value in the history. When the threshold increases, smaller differences between the spikes and the maximum are no longer detected. A bigger threshold excludes the smaller differences; this results in a smaller number of correctly detected shot boundaries.

A drop of the number of correct detections with an increased amount of Gaussian noise is probably caused by that the added Gaussian noise increases the amount of differences in the sequence. This increases the maximum value in the history. This, in turn decreases the difference between the spike and the maximum. That could result in a failure to detect the shot boundary.

In the plots of the false detection measurements (figures 5.21b, 5.22b, 5.23b and 5.24b), we observe a decrease of the number of false detections with an increase of the threshold. The detector with a low threshold can interpret increases in the difference measurement, caused by regular motion in the sequence, as a shot boundary more easily, than with a high threshold.

Increasing the amount of added Gaussian noise decreases the amount of false detections. The added Gaussian noise increases the maximum of the difference history. This decreases the difference between the spike and the maximum. This reduces the number of false detections.

Conclusions

Overall the following conclusions can be drawn:

- Our solution can detect hard cut shot boundaries.
- The higher the threshold is set, the smaller the probability for a false detection is, but also the smaller the chance for a correct detection is.
- The higher the level of the added Gaussian noise, the smaller the chance for a false detection, but also the smaller the chance for a correct detection.

We do not evaluate the performance of the hard-cut shot boundary detector, because:

- We do not have another detector to evaluate against.
- We currently do not have an application to test this detector.

Chapter 6

Evaluation

In this chapter we evaluate the improvements in cost, expressed in the amount of processor cycles spent and performance, expressed in the percentage fields of which the film mode is correctly detected. The improved Arrow difference metric is compared to the original difference metric of the SRFD.

6.1 Performance measurement

For the performance measurement the output of the original metric is compared with the output of the arrow metric. We compare the typical sequences and the problem sequences.

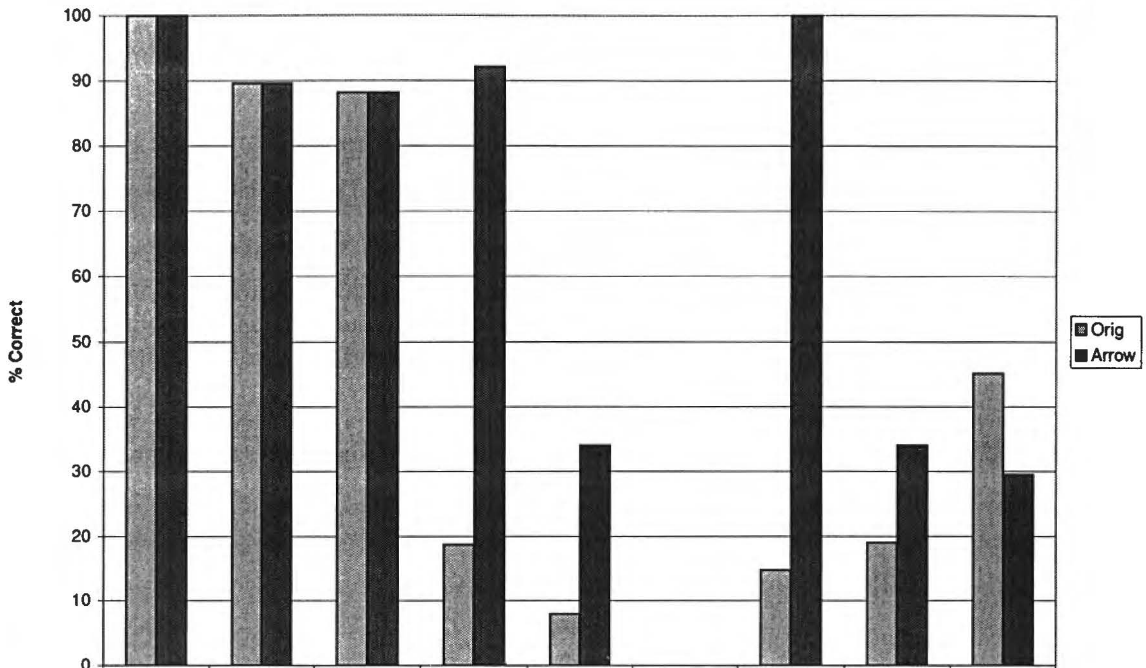
For all the measurements, the output of the field difference path is used. The output of the frame difference path is ignored, since the implementation of the Arrow based SRFD only differs in the filed difference film detection path.

When we compare the percentages in table 6.1 (figure 6.1), we observe that for the three typical sequences, Renata, Flight and Usual Suspects, both the difference metrics show almost identical results. During the run-in period, the film detector produces video mode. Video mode is the fall-back mode of our scan-rate converter. In the fall-back mode the scan-rate converter is the mode, in which the video quality of the signal produced by the scan-rate converter, is still of acceptable quality. The mode of the fields processed during this run-in is unknown, thus the film detector detects the fall-back mode, in this case video mode.

We can observe a large increase in percentage of the number of correctly detected fields in the Fargo Office sequence and an increase in the correctly identified fields in the Fargo Iceplane sequence. The RTL-z sequence is correctly classified as video mode for the whole sequence by the Arrow detector. The Ski Text sequence is still completely detected as a sequence in video mode, which is incorrect.

	Absdif				Arrow			
Sequence	#Vid	#F22	#F32	%Correct	#Vid	#F22	#F32	%Correct
Renata	200	0	0	100%	200	0	0	100%
Flight	15	129	0	89,6%	15	129	0	89,6%
Usual Suspects	33	0	255	88,2%	34	0	254	88,2%
Fargo Office	174	40	0	18,7%	17	197	0	92,1%
Fargo Iceplane	184	16	0	8%	132	68	0	34%
Ski Text	100	0	0	0%	100	0	0	0%
RTL-z	15	87	0	14,7%	102	0	0	100%
Parade	18	22	0	19%	16	24	0	34%
TMF	110	134	0	45,1%	72	172	0	29,5%

Table 6.1: Performance measurements on the original SRFD difference metric and the Arrow difference metric.



	Renata	Flight	Usual Suspects	Fargo Office	Fargo Iceplane	Ski Text	RTL-z	Parade	TMF
Orig	100	89.6	88.2	18.7	8	0	14.7	19	45.1
Arrow	100	89.6	88.2	92.1	34	0	100	34	29.5

Figure 6.1: Performance measurements on the original SRFD difference metric and the Arrow difference metric.

The results for the Parade and TMF sequences have to be interpreted carefully. The term correct indicates the percentage of the fields correctly identified. This correct identification however is dependent on the application of the film detector. As a detector that has to detect the presence of motion in pull-down mode, the film detector is correct to classify the TMF and Parade sequences as 2:2 pull-down film mode. The sequence contains motion that behaves as being processed with a 2:2 pull-down process. For the application of a film detector for advanced scan-rate converters, the correct classification of these sequences would be video mode. Since this is the fall-back mode of the scan-rate converter. Annoying artifacts are introduced if the elements that are in video mode are processed in film mode by the scan-rate converter. The application asks that, if video mode elements are present, the processing be done in a fall-back mode. We conclude from the measurements on both the TMF as the Parade sequences that neither detection scheme results in a more correct classification of the sequences.

By classifying these sequences as hybrid film mode, both application fields can be satisfied. In advanced scan-rate conversion, the hybrid mode sequences can be correctly processed in the fall-back mode. Using this classification (extending it with the hybrid mode) the results measured for the TMF and the Parade sequences are 0%, for both the original as the Arrow difference detector. This is not surprising, as the SRFD is unable to detect hybrid mode sequences.

6.2 Algorithmic cost

The algorithmic cost is measured using the TMSIM Tri-Media simulator. The cost is measured in number of processor cycles. However, despite of being a processor cycle accurate simulator, its configuration differs from hardware configuration, e.g. the access time to RAM is different in the hardware configuration. The numbers obtained from the simulations do not reflect the exact number of cycles spent in the hardware configuration. They do give a good indication of the cost and can serve as an accurate indicator of the increase in cost of a new algorithm. Therefore, we will use the numbers to estimate the increase of the cost

	full	sub-sampled
original	26.588.636	6.904.436
arrow	28.931.514	7.491.112

Table 6.2: Number of cycles spent on 50 fields (equals 1 second).

of the Arrow algorithm.

Each algorithmic run processes a sequence of 50 fields (being equivalent with 1 second of a 50 Hz video sequence). For both the original as the arrow algorithm, the cost is measured on a full search and a sub-sampled version. This sub-sampling is done applying the metric on every fourth line.

The result of these simulations are summarized in table 6.2.

We can observe that there is an increase of 8.5% in the amount of cycles spent, when we compare the Arrow algorithm to the original algorithm. The increase in cost does not cause the metric to exceed the budget of $10 \cdot 10^6$ cycles.

6.3 Conclusions

From these measurement we can conclude:

- The arrow difference detector does not significantly change the performance on the typical sequences.
- The arrow difference detector improves the performance on the Fargo Office and RTL-z sequences and to some degree the Fargo Iceplane sequence. This is the result of the insensitivity of to static vertical detail.
- The arrow difference detector increases the processor load with 8,5%. The total processor load of the sub-sampled version is below the $10 \cdot 10^6$ processor cycles per second limit.
- Neither of the algorithms can detect the film mode in the Fargo Ice Plane, Parade, TMF and Ski-text sequences accurately. In cases of the Fargo Iceplane and Ski Text sequences, the amount of noise seems to interfere with the correct detection of the film mode. The Parade and TMF sequences cannot be correctly classified by the SRFD. A new type of film detector is required to correctly classify these sequences.

Chapter 7

Overall Conclusions and Recommendations

7.1 Overall Conclusions

Two goals were defined for this project:

- gaining an overview of the field of film detection technology.
- improving the performance of the SRFD without an significant increase in algorithmic cost.

For both goals, results were obtained. An overview of film detector technology, based on the available literature, was created. From this overview, it became apparent that films detectors exist that are able to detect both 2:2 pull-down and 3:2 pull-down film modes. These film detectors are used in the field of frame-rate conversion and encoding for compression. For 3:2 pull down film mode, a robust and economic solution is known. The challenge lies at robust and cheap 2:2 pull-down film mode detection. The interlacing of the video signal causes the detection to be less robust, than with 3:2 pull-down detection. The major problems that were mentioned in the literature were: False motion detection, video composites and Video edits.

The Schutten-Riemens Film Detector (SRFD) was used for further research. Using the SRFD, several sequences were identified that are incorrectly detected by the SRFD. These sequences were characterized by: noise, false detection due to vertical detail and containing mixed mode. In particular, the sequence (Parade) was identified. In this sequence, the mixed mode was probably caused by the DV-encoding scheme used on that signal. This is the first time that this method is identified as a source for mixed mode signals.

A new classification is required for mixed mode signals. We introduce the Hybrid mode, to correctly classify these mixed modes.

For reducing the effect of false detection due to static vertical detail, the Arrow difference metric is tested. Using this metric, the SRFD correctly detects the film mode of the problem sequences associated with this problem. The solution increased the cost of the implementation with 8,5%. The total cost, using a sub-sampling scheme, remained below the limit of 10^7 processor cycles per second.

As a second improvement to the SRFD, a hard cut shot boundary detector is implemented and tested. The detector is able to identify hard cut shot boundaries. The behavior of the hard cut shot boundary detector with a varying Gaussian threshold and varying noise levels has been tested. An increased threshold leads into fewer shot boundaries to be detected. However, the number of incorrectly identified shot boundaries reduces as well. An increased Gaussian noise level also leads to a reduction in the number of correctly identified shot boundaries. The number of incorrectly identified shot boundaries drop as well with an increasing Gaussian noise level.

Several of options for improvement remain. These are robustness against video edits, hybrid mode detection and robustness against noise.

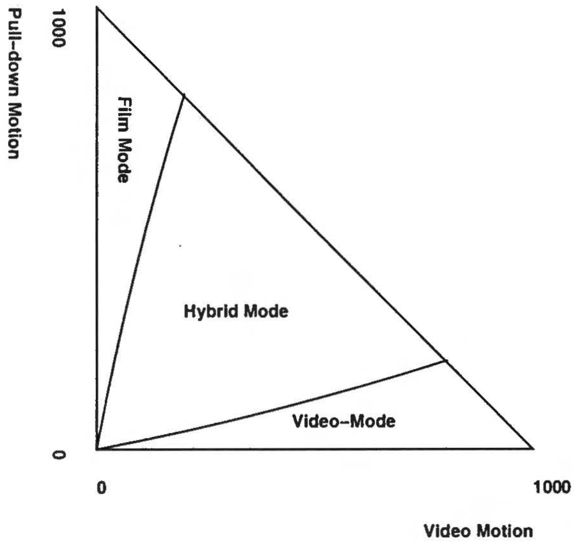


Figure 7.1: Example classification using two motion metrics.

7.2 Recommendation I: Dual motion

Traditionally, two classes of film modes are defined: film mode and video mode. However instances have been observed, as the Parade and TMF sequences, where both pull-down and video motion patterns occur. These sequences do not fit in the traditional classification that has been reported in the prior art. The SRFD and other film detectors fail to robustly classify these sequences, because they are based on this traditional classification.

This can cause problems when processing these sequences. For current advanced scan rate conversion, the sequences should be treated using a fall-back strategy. Since the SRFD cannot guarantee robust detection of these sequences, such a fall-back strategy cannot be implemented reliably.

In order to correctly process these sequences we propose a new classification scheme. This classification consists of two estimates: One estimate of the area of the picture that exhibits film motion patterns and one estimate of the area of the picture that exhibits video motion patterns.

These two estimates can be formulated as:

Pull-down Motion This is a motion pattern, where the motion behaves as in a film sequence. For 2:2 pull-down, this pattern will be an alternating pattern of motion and no-motion.

Video Motion This is a motion pattern, where the motion is continuous. There is no alternating pattern of motion and no-motion.

These two metrics occupy a position in the dual motion space for each field in a sequence. This is illustrated in figure 7.1

Ideally, sequences behavior in the dual motion space should look like this:

Video Mode All points occupied by this sequence should lie along the video motion axis.

Film Mode All points occupied by this sequence should lie along the film motion axis.

Hybrid Film Mode All points occupied by this sequence should not lie along the film motion axis or the video motion axis.

An implementation of this scheme involves three motion numbers that are estimated. Besides an estimate for the amount of video motion, there are also estimates for both phases of the pull-down process. By taking the absolute difference of these two pull-down phases, a film mode motion estimate can be obtained. By comparing the two film motion estimates, a phase can be extracted from these estimates.

Curr	Prev	Detected	Expected
A	A	L	L
B	A	H	H
B	B	L	L
C	B	H	H
C	C	L	L
D	C	H	H
D	D	L	L
E	D	H	H
E	E	L	L

a (edit between rows C C and D C)

Curr	Prev	Detected	Expected
A	A	L	L
B	A	H	H
B	B	L	L
C	B	H	H
D	C	**H**	**L**
E	D	H	H
E	E	L	L
F	E	H	H
F	F	L	L

b (edit between rows C B and D C)

Curr	Prev	Detected	Expected
A	A	L	L
B	A	H	H
B	B	L	L
C	B	H	H
D	C	**H**	**L**
D	D	**L**	**H**
E	D	**H**	**L**
E	E	**L**	**H**
F	E	**H**	**L**
F	F	**L**	**H**

c (edit between rows C B and D C)

Table 7.1: Three different types of video edits. A correct video edits (a), an incorrect video edit without a pattern break (b) and an incorrect video edit with a pattern break (c). The horizontal line indicates the moment of edit. The bold values are inconsistent with the expected pattern.

7.3 Recommendation II: Shot Boundary Based Video Edit Detection

In the prior art, video edits are considered a problem in the field of film detection. Video edits are caused by incorrect editing of film material after the transfer to video. This editing causes a discontinuity in the difference patterns, causing a potential incorrect detection of the film mode around the edit.

For hard edits in 2:2 pull-down film mode, three situations can occur.

Correct video edit Two film sequences are correctly edited together. In this case, the edit has the same properties as an edit in the film material before it has been transferred to video. This kind of video edit does not cause problems for film detectors. This type is shown in figure 7.1. The horizontal line indicates the edit moment. Each frame is repeated the correct amount. There is no difference between the detected and the expected HL-values.

Incorrect video edit without pattern break Two film sequences are edited together. The second sequence continues the phase of the first sequence. However, the shot boundary causes the last field not to be repeated. The repeat is replaced with a repeat from the second sequence. This replacement results in a H value from the difference detector, where a L value is expected. This type of edit is shown in figure 7.1b. This H-value is the only value that differs with the regular pattern. By detecting a hard-cut shot boundary at this position, a scheme can be devised where the detector ignores that different H-value. This would prevent the film detector to detect an erroneous film mode.

Incorrect video edit with pattern break Two film sequences are edited together. The phases of the sequences differ. All H and L-values after the edit are detected incorrectly. This type of edit is shown in figure 7.1c. All HL-values after the video edit differ from the expected values.

In order to correctly process material containing incorrect video edits, we need to distinguish an extra measurement that can determine if an incorrect value is caused by a video edit.

We propose to create a solution based on the following assumption:

Video edits occur at shot boundaries.

Using a reliable shot detector, we can create a scheme that can detect potential video edits by detecting shot boundaries. Such a scheme can potentially prevent the incorrect detection of the film mode and prevent a new run in period after a video edit.

An implementation using information created by a shot detector should improve the performance of the film detector at video edit. Therefore, we recommend to research such a scheme.

Bibliography

[1] R. Armitano. Low-cost telecine detection for real-time video coding. *Proceedings-of-the-SPIE*, 3528:261–268, 1999.

[2] E.B. Bellers. *De-interlacing: A contribution to the interlaced versus progressive video debate*. PhD thesis, Technische Universiteit Delft, 2000. ISBN 90-74445-47-0.

[3] M. A. Bock. Removal of redundant field in standards conversion. Applicant: Digi-Media Vision Limited, London, UK, April 9, 1996. London: Patent Office, UK Patent Application GB 2312116A.

[4] S.D. Casavant, R.N. Hurst, S.S Perlman, M.A. Isnardi, and F. Aschwanden. Video/film-mode (3:2 pulldown) detector using patterns of two-field differences. Assignee: RCA Thomson Licensing Corporation, Princeton, N.J., May 31, 1994. United States Patent Office US 5,317,398.

[5] T.J. Christopher and C. Correa. Method and apparatus for identifying video fields produced by film sources. Assignee: Thomson Consumer Electronics, Inc., Indianapolis, Ind., November 18, 1997. United States Patent Office US 5,689,301.

[6] T.J. Christopher and C. Correa. Method and apparatus for identifying video fields from film sources employing 2-2 and 3-2 pull down sequences. Assignee: Thomson Consumer Electronics, Inc., Indianapolis, Ind., January 26, 2000. European Patent Office, European Patent EP 0720366B1.

[7] T.J. Christopher and C. Correa. Method and apparatus for identifying video fields produced by film sources. Assignee: Thomson Consumer Electronics, Inc., Indianapolis, Ind., February 9, 2000. European Patent Office, European Patent EP 0720367B1.

[8] G.R. Coombs, D. Andrew, and O.J. Morris. Identifying film frames in a video sequence. Assignee: Philips Electronics UK Limited, Croydon, GB, December 22, 1999. European Patent Office, European Patent EP 0612187B1.

[9] C. Correa and R. Schweer. Method and device for film-mode detection. Assignee: Deutsche Thomson-Brandt GmbH, Villengen-Schwenningen, Germany, November 15, 1994. United States Patent Office US 5,365,273.

[10] 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, European Patent EP 0567072B1.

[11] S. Del Corso. Film source video detection and encoding. Applicant: Koninklijke Philips Electronics N.V., Eindhoven, NL, September 7, 2001. World intellectual property organization, International Publication Number: WO 01/65857 A1.

[12] R.A. Dischert. Progressive scan display system emplying line and frame memories. Assignee: RCA Corporation, Princeton, N.J., US, February 3, 1987. United States Patent Office US 4,641,188.

[13] Y.C Faroudja. Film-to-video converter with scan line doubling. Assignee: Faroudja, Y.C., Los Altos, Calif, US, October 24, 1989. United States Patent Office US 4,876,596.

[14] Y.C. Faroudja, P.D. Swartz, and J.J. Campbell. High-definition television signal processing for transmitting and receiving a television signal in a manner comaptible with the present system. Assignee: Farjoudja Laboratoies, Inc., Sunnyvale, Calif., August 22, 2000. United States Patent Office US 6,108,041.

[15] B.A. Flannaghan. Apparatus for processing a television signal including a movement detector. Assignee: Independent Broadcasting Authority, England, GB, October 27, 1987. United States Patent Office US 4,703,358.

[16] P. Gerets. Method and apparatus for video signal processing. Applicant: Barco N.V., Poperlinge, Be, March 23, 2000. World intellectual property organization, International Publication Number: WO 00/16561.

[17] G. de Haan. *Video Processing for multimedia systems*. University Press Eindhoven, 2000. ISBN 90-9015015-8.

[18] G. de Haan, P.W.A.C. Biezen, and O.A. Ojo. An evolutionary architecture for motion-compensated 100 hz television. *IEEE transactions on circuits and systems for video technology*, 5(3):207–217, June 1995.

[19] G. de Haan, H. Huijgen, P.W.A.C. Biezen, and O.A. 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, N.Y., November 15, 1994. United States Patent Office US 5,365,280.

[20] A. Hanjalic. Shot-boundary detection: unraveled or resolved? *IEEE Transactions on Circuits and Systems for Video Technology*, 12(2):90–105, February 1991.

[21] ITU Radiocommunication Assembly. *Conventional television systems*, 1998. Rec. ITU-R BT.470-6.

[22] ITU Telecommunications Standardization Sector. *Information technology - Generic coding of moving pictures and associated video information: video*, 1995. Rec. ITU-T H.262.

[23] E.A. Krause. Progressive scan display of video derived from film. Assignee: General Instrument Corporation, New York, N.Y., US, November 14, 1989. United States Patent Office US 4,881,125.

[24] Young Sun Lee, Jin Hwan Lee, Yo Sung Ho, and Joo Hong Jeong. Film mode video sequence detector. Assignee: Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, March 31, 1998. United States Patent Office US 5,734,420.

[25] Il Taek Lim. Film mode detection method using periodic pattern of video sequence. Assignee: LG Electronics Inc. Seoul, KR, December 1, 2000. United States Patent Office US 2001/0002853.

[26] Hui Yau Wei Lucas. Progressive/interlace and redundant field detection for encoder. Applicant: STMICRO-Electronics Asia Pacific PTE LTD, Singapore, June 8, 2000. World intellectual property organization, International Publication Number: WO 00/33579.

[27] G. Lupatini, C. Saraceno, and R. Leonardi. Scene break detection: a comparison. *Research Issues In Data Engeneering*, pages 34–41, 1998. Proceedings of 'Continuous-Media Databases and Applications'.

[28] J.J. Lyon, T.C.AND Campbell. Motion sequence pattern detector for video. Assignee: Faroudja, Yves C., Los Altos Hills, CA, US, January 1, 1991. United States Patent Office US 4,982,290.

[29] A. Martin. Method and device for film-mode detection. Assignee: Thomson Consumer Electronics, Inc., Indianapolis, Ind., April 11, 1995. United States Patent Office US 5,406,333.

[30] A. Martin and M. Smith. Method and device for film-mode detection and elimination. Assignee: Thomson Consumer Electronics, Inc., Indianapolis, Ind., September 19, 1994. United States Patent Office US 5,452,011.

[31] J. van der Meer and F.W.P. Vreeswijk. Method of and arrangement for motion detection in an interlaced television picture obtained after film-to-television conversion. Assignee: U.S. Philips Corporation, New York, N.Y., US, June 12, 1990. United States Patent Office US 4,933,759.

[32] Chong-Wah Ngo, Ting-Chuen Pong, Hong-Jiang Zhang, and R.T. Chin. Motion-based video representation for scene change detection. *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 1:827–830, 2000.

[33] K. Otsuji, Y. Tonomura, and Y. Ohba. Video browsing using brightness data. *Proceedings of the SPIE The International Society for Optical Engineering (USA)*, 1606:980–9, 1991.

[34] Philips Semiconductors. *Data Sheet SAA4990H Progressive scan-Zoom and Noise reduction IC (PROZONIC)*, October 25, 1996.

[35] Philips Semiconductors. *Data Sheet SAA4993H Field and line rate converter with noise reduction (FALCON)*, November 23, 2001.

[36] K.H. Powers. High definition television signal for film-television standards conversion system. Assignee: RCA Corporation, Princeton, N.J., US, December 30, 1986. United States Patent Office US 4,633,293.

[37] S. Rathnam and G. Slavenburg. An architectureal overview of the programmable multimedia processor, TM-1. *Compcon '96. 'Technologies for the Information Superhighway' Digest of Papers*, pages 319–326, 1996.

[38] A.K. Riemens, R.J. Schutten, and K.A. Vissers. High speed video de-interlacing with a programmable trimedia vliw core. *Proceedings of the International Conference on Signal Processing Applications and Technology (ICSPAT)*, pages 1375–1380, 1997.

[39] R.J. Schutten and A.K. Riemens. Condor video scan rate conversion, part 2: Algorithms. Nat.Lab. Report 7222, Koninklijke Philips Electronics N.V., March 2, 2002.

[40] P.L. Swan. System and method for reconstructing noninterlaced captured content for display on a progressive screen. Assignee: ATI Technologies, Inc., Thornhill, Canada, April 25, 2000. United States Patent Office US 6,055,018.

[41] P.D. Swartz. Film source video detection. Applicant: Faroudja Laboratories, Inc., Synnyvale, Ca, US, April 22, 1999. World intellectual property organization, International Publication Number: WO 99/20040.

[42] P.D. Swartz. Film source video detection. Assignee: Faroudja Laboratories, Inc., Sunnyvale, Ca, US, January 11, 2000. United States Patent Office US 6,014,182.

[43] P.D. Swartz. Film source video detection. Assignee: Faroudja Laboratories, Inc., Sunnyvale, CA (US), March 13, 2001. United States Patent Office US 6,201,577 B1.

[44] Lu Tong and P.N. Sugantan. An adaptive cumulation algorithm for video shot detection. *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing*, pages 296–299, 2001.

[45] Kazuro Washi and Masamori Oguino. Television system. Assignee: Hitachi, Ltd., Japan, January 19, 1988. United States Patent Office US 4,720,744.

[46] Yoichi Yagasaki and Teruhiko Suzuki. Coding and decoding digital video signals having duplicate pictures and frames with fields originating from different film source frames. Assignee: Sony Corporation, Tokyo, Japan, June 9, 1999. European Patent Office, European Patent EP 0588668B1.

[47] Hirohisa Yamaguchi, Masahiro Wada, and Hideo Yamamoto. Frame rate conversion system in television signal. Assignee: Kokusai Denshin Denwa Co., Ltd., Tokyo, Japan, June 9, 1987. United States Patent Office US 4,672,442.

[48] Jun Yonemitsu, Teruhiko Suzuki, and Yoichi Yagasaki. Apparatus for coding and decoding a digital video signal derived from a motion picture film source. Assignee: Sony Corporation, Japan, October 24, 1995. United States Patent Office US 5,461,420.

Appendix A

Literature Research

A.1 Summary final project

Philips has a strong position in the market of video format conversion for consumer electronics (known as 'natural motion' in TV sets). In order to achieve high image quality, motion estimation and motion compensation techniques are applied. A format converter is used, that consists of functions as film detection, motion estimation, de-interlacing and image interpolation. Currently, Philips products use a hardware-based implementation of the format converter. In a project, Philips Research has developed novel algorithms and an innovative architecture for the next generation format converters, aiming at improved image quality and increased flexibility. For consumer electronics applications, the cost of the system is an extremely important design constraint, posing severe challenges throughout the research, development and implementation trajectory. The format converter consists of a dedicated hardware block for pixel rate operations and software running on a powerful embedded CPU for image and block rate operations, in contrast to the current pure hardware implementation.

As mentioned one of the functions of the format converter is film detection. Film detection, or also known as telecine detection is a function that is implemented in software. This function attempts to determine whether an original video source is a movie camera or a video camera. A movie camera records 24 images per second on film, while a video camera records 50 (or 60 depending video format used) images per second. A signal originating from a film camera is converted during transfer to video 50 (or 60) images per second by repeating the most recent image. This operation has significant impact on the smoothness of motion in the video sequence. The transferred film sequence shows jerky motion. This jerky motion is called motion judder.

A film detector analyses the video signal to distinguish if the original source was film mode or video mode. All other functions in a scan rate conversion system depend on reliable detection of this film mode, since video that originates from film requires different video processing than video that does not originate from film. Video sequences originating from film need removal of motion judder, while a pure video sequence this removal is undesired.

During the first part of this assignment, I will investigate options to reduce the CPU load of the film detector, preserving as much image quality as possible. If it is possible to reduce the CPU load sufficiently, the implementation can be cheaper by using a cheaper CPU and thus the market potential of the format converter will increase. Based on an existing ANSI-C model of the hardware format converter, various potential solutions are implemented and the results judged using advanced video simulation tools. In the second part, the current software algorithm can be improved and then implemented on real-time hardware to assess cost effectiveness of the improvements.

A.2 Literature Research Assignment

The technique of motion compensated frame rate conversion has only recently become feasible. This technique requires a film detector to reproduce the original film sequence. Because of the novelty of this technique, it is expected that solutions have emerged only recently. It is however, important to find 'proven technology'. Since film detectors are applied in commercial products, the literature research should include a patent search. Interest goes out to specialized literature, dealing with algorithms and implementations of film detectors.

The goal of this literature research is to obtain an overview of the field of film or telecine detection. Telecine conversion is practiced from the first television broadcasts. Detection of this conversion started around 1990. The scope of this literature research is the period from 1991 up to now.

Only a few institutes are engaged in this field of research, so the scope of the search should be global. Television manufacturers and MPEG encoding are known to apply film detectors.

A.3 Conceptual Index of Final Report

The concept index of the project's final report is:

1. Introduction
2. Current State of Technology
3. The SRFD (Schutten-Riemens Film Detector)
4. Improvements to the SRFD
5. Evaluation, Conclusions and Recommendation

This literature search forms the basis for chapter 2. In that chapter, the current state of technology is explained.

A.4 Terms used

Based on private discussions with Prof. de Haan and Ing. A.K. Riemens, initial search terms were formed. These terms are

1 film detection

2 film detector

3 frame rate detection

4 frame rate detector

5 movie detection

6 movie detector

7 telecine

These terms are based on the prior knowledge of film detectors at Philips Research.

A.5 List of sources an the number of initially selected literature references.

The literature research started with the search engine of the Eindhoven University of Technology (TU\e), VUBIS. Using 'word from title' this resulted for search terms 1-7 in no hits. This indicates, as expected, that the problem is highly specialized.

Next, the INSPEC reference database was searched using the search terms 1 through 7. This resulted in the following number of hits:

Term	Number of hits
1	33
2	227
3	5
4	1
5	0
6	0
7	321

From the results of INSPEC, the poor choice of search terms became apparent. Term 1 and 2 are also used for the detection of, for example, a film of oil on water. Terms 3 and 4 are also used commonly in telecommunications. Term 7 indicates the machine to transfer film to video. This led to a refinement of the search terms for INSPEC.

8 Item 2 is combined with the word video to limit the number of hits.

9 Item 7 is combined with the term 'detection'.

10 Item 7 is combined with the term 'detector'.

INSPEC resulted in:

Term	Number of hits
8	1
9	9
10	3

Of these hits, only publication that covered the subject of film detection in video was by Armitano [1].

A.6 Selection Criteria

Due to the apparent specialty and novelty of the subject, the only publication found on this subject is by Armitano [1]. Other papers cover the field of frame rate conversion or de-interlacing. These fields rely on film detection, but do not cover the subject. Therefore, they do not qualify as sources for further research.

A.7 Snowball and citation method

The paper by Armitano [1], is the only paper that covers the area of the assignment. Therefore, this paper as the starting point for the snowball search. The results of this search are displayed in A.1. The publications quoted in figure A.1 are listed below in inverse chronological order.

- Armitano, R, LOW-COST TELECINE DETECTION FOR REAL-TIME VIDEO CODING,SPIE Vol. 3528, 1998 p. 261-268.
- Kato M., Oda T. and Tahara K., METHOD AND APPARATUS FOR ENCODING MOVING PICTURE SIGNALS AND RECORDING MEDIUM FOR RECORDING MOVING PICTURE SIGNALS, Applicant: Sony Corporation, Tokyo, Japan, Oct. 7, 1997, United States Patent 5,675,379
- Christopher T.J. and Correa C., METHOD AND APPARATUS FOR IDENTIFYING VIDEO FIELDS PRODUCED BY FILM SOURCES EMPLOYING 2-2 AND 3-2 PULL DOWN SEQUENCES, Applicant: Thomson Consumer Electronics, Inc., Indianapolis, Ind, USA, Oct 8. 1996, United States Patent 5,563,651
- Hewlett G.J. and Gove R.J., ENCODING DATA CONVERTED FROM FILM FORMAT FOR PROGRESSIVE DISPLAY, Applicant: Texas Instruments Incorporated, Dallas, Tex. USA, April 16, 1996, United States Patent 5,508,750
- Martin A. and Smith M., METHOD AND DEVICE FOR FILM-MODE DETECTION AND FIELD ELIMINATION, Applicant: Thomson Consumer Electronics, Inc., Indianapolis, Ind, USA, Sep. 19, 1995, United States Patent 5,452,011

- Martin A., METHOD AND DEVICE FOR FILM-MODE DETECTION, Applicant: Thomson Consumer Electronics, Inc., Indianapolis, Ind, USA, Apr. 11, 1995, United States Patent 5,406,333
- Casavant S.D., Hurst R.N., Perlman S.S., Isnardi M.A. and Aschwanden F., VIDEO/FILM-MODE (3:2 PULL-DOWN) DETECTOR USING PATTERNS OF TWO FIELD DIFFERENCES, Applicant: RCA Thomson Licensing Corporation, Princeton, N.J., USA, May 31, 1994, United States Patent 5,317,398
- Faroudja Y.C., Dong Xu and Swartz P., MOTION DETECTION BETWEEN EVEN AND ODD FIELDS WITHIN 2:1 INTERLACED TELEVISION STANDARD, Applicant: Faroudja Laboratories, Sunnyvale Ca., Mar. 1, 1994, United States Patent 5,291,280
- Weckenbrock H.J. and Christopher S.H., MOTION DETECTION FOR VIDEO INCLUDING THAT OBTAINED FROM FILM, Applicant: Samsung Electronics Co., Ltd., Sowun, Rep. of Korea., Nov. 30, 1993, United States Patent 5,267,035
- Richards J.W., Krsljanin M. and Ozaki Y., VIDEO POSTPRODUCTION OF MATERIAL ACQUIRED ON FILM, Applicant: Sony Broadcast & Communications Limited, Basingstoke, England, Mar. 2, 1993, United States Patent 5,191,427
- Ishii H. and Morimura A., IMAGE MOTION VECTOR DETECTING APPARATUS, Applicant: Matsushita Electric Industrial Co., Ltd., Osaka, Japan, May 5, 1992, United States Patent 5,111,511
- Lyon T.C. and Campbell J.J., MOTION SEQUENCE PATTERN DETECTOR FOR VIDEO, Applicant: Faroudja Y.C., Data of Patent: Jan 1, 1991, United States Patent 4,982,280
- Meer J. van der and Vreeswijk W.P., METHOD OF AND ARRANGEMENT FOR MOTION DETECTION IN AN INTERLACED TELEVISION PICTURE OBTAINED AFTER FILM-TO-TELEVISION CONVERSION, Applicant: U.S. Philips Corporation, New York, N.Y., USA, Jun. 12, 1990, United States Patent 4,933,759
- Krause E.A., PROGRESSIVE SCAN DISPLAY OF VIDEO DERIVED FROM FILM, Applicant: General Instrument Corporation, New York, N.Y., Nov. 14, 1989, United States Patent 4,881,125
- Faroudja Y.C., FILM-TO-VIDEO CONVERTER WITH SCAN LINE DOUBLING, Applicant: Faroudja Y.C., Oct. 24, 1989, United States Patent 4,876,596
- Washi K and Oguino M., TELEVISION SYSTEM, Applicant: Hitachi, Ltd., Japan, Jan. 19, 1988, United States Patent 4,720,744
- Flannaghan B.A., APPARATUS FOR PROCESSING A TELEVISION SIGNAL INCLUDING A MOVEMENT DETECTOR, Applicant: Independent Broadcasting Authority, England, Oct. 27, 1987, United States Patent 4,703,358
- Yamaguchi H. Wada M. and Yamamoto H., FRAME RATE CONVERSION SYSTEM IN TELEVISION SIGNAL, Applicant: Kokusai Denshin Denwa Co., Ltd., Tokyo, Japan., Jun. 9, 1987, United States Patent 4,672,442
- Dischert R.A., PROGRESSIVE SCAN DISPLAY SYSTEM EMPLOYING LINE AND FRAME MEMORIES, Applicant: RCA Corporation, Princeton, N.J., USA, Feb. 3, 1987, United States Patent 4,641,188
- Powers K.H., HIGH DEFINITION TELEVISION SIGNAL FOR FILM-TELEVISION STANDARDS CONVERSION SYSTEM, Applicant: RCA Corporation, Princeton, N.J., Dec. 30, 1986, United States Patent 4,633,293
- MassMann V., METHOD FOR THE TELEVISION SCANNING OF FILMS, Applicant: Bosch R., Stuttgart, Fed. Rep. of Germany, Aug. 24, 1982, United States Patent 4,336,408

- Poetsch D., METHOD AND APPARATUS FOR THE TELEVISION SCANNING OF FILMS, Applicant Bosch R., Stuttgart, Fed. Rep. of Germany, Jan. 12, 1982, United States Patent 4,310,856
- Michael P.C., Taylor R.J. and Kellar P.R.N., T.V. PICTURE FREEZE SYSTEM, Applicant: Micro Consultants Limited, Berkshire, England, Jun. 9, 1981, United States Patent 4,272,787
- Millward J.D., TELEVISION FILM SCANNER, Applicant: The Rank Organisation Limited, London, England, May 27, 1980, United States Patent 4,205,337
- Zinchuk M., APPARATUS AND METHOD FOR DISPLAYING MOVING FILM ON A TELEVISION RECEIVER, Applicant: Polaroid Corporations, Cambridge, Mass., USA, Apr. 24, 1979, United States Patent 4,151,560
- Longchamp J.F., METHOD AND SYSTEM FOR CONVERTING THE IMAGE CONTENT OF TRANSPORTED FILM ONTO TELEVISION SIGNAL PICTURE INFORMATION, Applicant: Longchamp J.F., Apr. 10, 1979, United States Patent 4,149,191
- Biber C.H., ELECTRONIC SOUND MOTION PICTURE PROJECTOR AND TELEVISION RECEIVER, Applicant: Polaroid Corporation, Cambridge, Mass., USA, Apr. 27, 1976, United States Patent 3,953,885
- Sanderson R.L. and Harrison M.G., OPTICAL-TO-ELECTRICAL SIGNAL TRANSDUCER METHOD AND APPARATUS, Applicant: Eastman Kodak Company, Rochester, N.Y., USA, Apr. 9, 1974, United States Patent 3,803,353
- Gold N., Ting L.K.M. and Weeks R.F., COLOR TELEVISION SYSTEM, Applicant: Polaroid Corporation, Cambridge, Mass., USA, Apr. 14, 1970, United States Patent 3,506,778

The citation method started with an European patent handed to me by Prof.dr.ir. G. de Haan. The results of the snowball method cover the origins of film detectors, but barely covers film detectors for European video systems. Therefore European patent EP 0 567 072 is the start of the citation search. Some citations move downwards, citing in the future. This is caused by the fact that the date of the application is recorded differently than the date of the patent. With the application the date of application counted. With the patent the 'Date of patent' is counted. Furthermore, the 'Date of Patent' does not strictly represent the chronological order, since the time to process the patent application can vary.

The patents quoted in figure A.2 are listed below in chronological order, application on date of application, patents on date of patent.

- Correa, Carlos and Schweer, Rainer, Verfahren und Vorrichtung zur Film-Mode-Detektion, Patentinhalber: DEUTSCHE THOMSON-BRANDT GMBH, Villingen-Schwenningen, DE, 1 june 1998, European Patent 0 567 072 B1
- Hackett, A, METHOD AND APPARATUS FOR REDUCING CONVERSION ARTIFACTS, Assignee: Thomson Consumer Electronics S.A., Courbevie, France, March 11 1997 United States Patent 5,610,662
- Hackett, A, Method and apparatus for reducing conversion artefacts, Applicant: THOMSON CONSUMER ELECTRONICS (S.A.), Courbevoie (FR), October 4 1995, European Patent Application 0 675 643 A1
- Faroudja, Y.C. and Xy, Dong, TELEVISION SIGNAL PROCESSING APPARATUS INCORPORATING MEANS FOR DETECTING 25 FRAME/SECOND MOTION PICTURE FILM SOURCES IN 50 Hz TELEVISION SIGNALS, Applicant: Faroudja, Y.C., Los Altos Hills, Ca, US, December 22 1994, INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT) WO 94/30006
- Tani, Masahiro and Okumara, Naoji, Detection of average luminance levels in different areas of a video image, and automatic discrimination of the image display format based here on, Assignee: MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD. Osaka, JP., Jun 6 2001, European Patent Specification EP 0 716 542 B1

- Swartz, P.D., FILM SOURCE VIDEO DETECTION, Assignee: Faroudja Laboratories, Inc., Sunnyvale, Ca, US., March 13, 2001, United States Patent 6,201,577 B1

- Faroudja, Y.C., Swartz, P.D, and Campbell, J.J., HIGH-DEFINITION TELEVISION SIGNAL PROCESSING FOR TRANSMITTING AND RECEIVING A TELEVISION SIGNAL IN A MANNER COMPATIBLE WITH THE PRESENT SYSTEM, Assignee: Faroudja Laboratories, Inc., Sunnyvale, Ca, US., August 22, 2000, United States Patent US 6,108,041

- Swartz, P.D., FILM SOURCE VIDEO DETECTION, Assignee: Faroudja Laboratories, Inc. Sunnyvale, Ca, US, January 11, 2000, United States Patent 6,014,182

- Heimburger, C., PROCESS FOR CORRECTION AND ESTIMATION OF MOVEMENT IN FRAMES HAVING PERIODIC STRUCTURES, Assignee: Thomson multimedia S.A., France, March 30, 1999, United States Patent 5,889,890

- Western, L.A., LOW-DELAY CONVERSION OF 3:2 PULL-DOWN VIDEO TO PROGRESSIVE FORMAT WITH FIELD AVERAGING, Assignee: Sharp Laboratories of America, Inc. Camas, WA, US., October 2, 2001, United States Patent 6,297,848

- Wagner, P., Schendowius, J., Zimmerman, K. and Erdler, O., WEIGHTED MEDIAN FILTER INTERPOLATOR, Assignee: Sony International (Europe) GmbH, Koln, De, April 17, 2001, United States Patent 6,219,102

- Weitbruch, S., Method and apparatus for automatic format detection in digital video pictures, Applicant: DEUTSCHE THOMSON-BRANDT GMBH, Villingen-Schwenningen, DE, May 6, 1999, European Patent Application EP 0 913 994 A1

- Conklin, G., SYSTEM AND METHOD FOR GENERATING VIDEO FRAMES AND DETECTING TEXT, Applicant: REAL-NETWORKS, INC. Seattle, US, January 4, 2001, INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT) WO 01/01701 A1

- Hirano Yasuhiro, Ishikaru Kazuo, Sugiyama Masato, Nakajima Mitsuo, Kurita Toshiyuki, Takata Haruki, Kimura Shoji, Tsuru Yasutaka, Kanehachi Takashi and Matono Takaaki, Conversion apparatus for image signals and TV receiver, Applicant: Hitachi, Ltd. Tokyo, JP, December 9, 1998, European Patent Application EP 0 883 298 A2

- Konishi Kazuo, Satoh Kohichi and Akamatsu Naoki, Letter image detection apparatus, Applicant KABUSHIKI KAISHA TOSHIBA, Kanagawa-ken, JP, April 22, 1998, European Patent Application, EP 0 837 602 A2

- Joanblanq, A., Process for detecting black bars in a video image, Applicant: THOMSON multimedia, Boulonge Billancourt, Fr., November 8, 2000, European Patent Application EP 1 051 033 A1

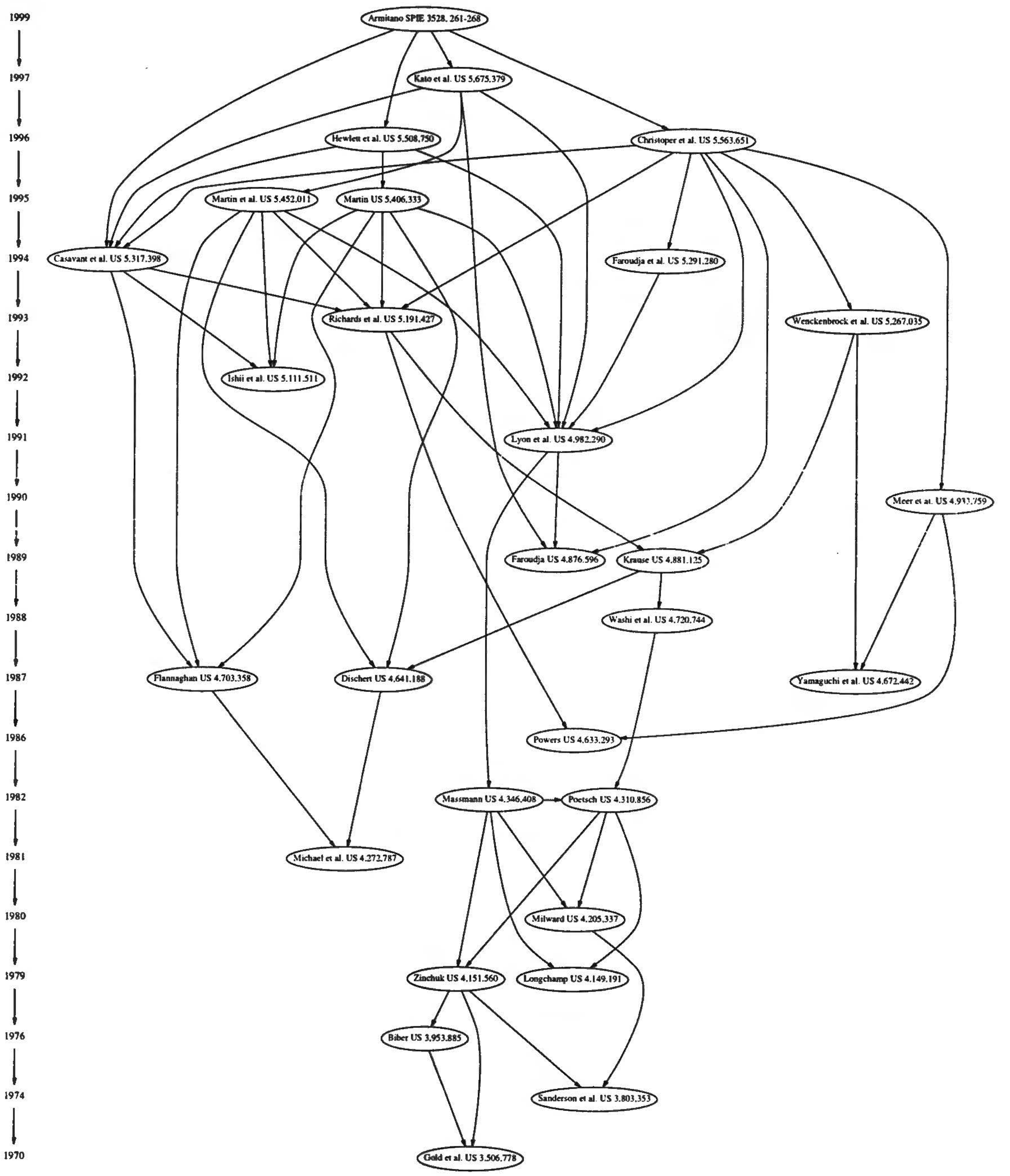


Figure A.1: Snowball method diagram.

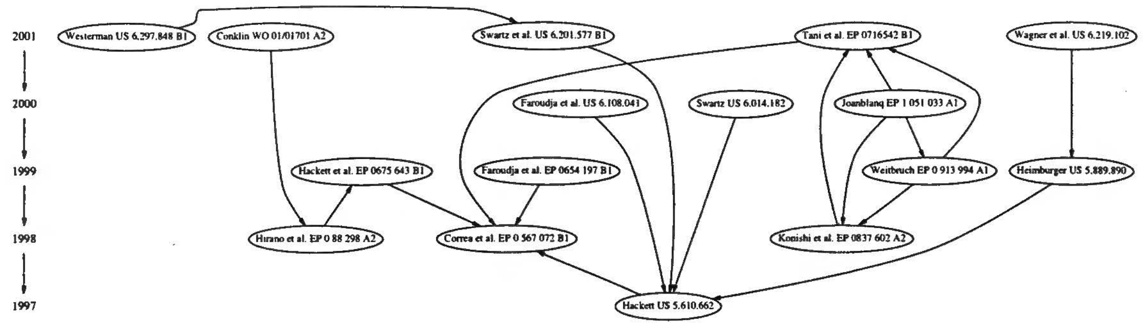


Figure A.2: Citation method diagram.

A.8 Relational pattern

	Chapter1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
1		X				
2		X		X		
3		X				
4		X				
5		X				
6		X				
7		X				
8		X				
9		X				
10		X				
11		X				
12		X				
13		X				
14		X				
15		X				
16		X				
17	X	X				
18	X					
19		X	X			
20				X		
21		X				
22		X				
23		X				
24		X				
25		X		X		
26				X		
27		X				
28		X				
29		X				
30		X				
31				X		
32				X		
33	X	X				
34	X		X			
35		X				
36			X			
37		X				
38		X	X	X		
39		X				
40		X				
41				X		
42		X				
43		X				
44		X				
45		X				

A.9 Conclusions

From the literature search itself and the resulting documents, the following facts can be observed.

- The problem field of film detection is highly specialized and related to the field of frame rate conversion, in particular, the field of de-interlacing and the field of video encoding.
- Almost every publication is a patent. This indicates that the subject is mainly handled in industrial research.
- The most important patent is by Lyon and Campbell[28]. Its claims cover the basic structure of film detectors. The assignee of this patent is Faroudja Laboratories. They are, together with Thompson Electronics, the most important publishers of patents in the field of film detectors.

- The proposed film detectors all use an interlaced luminance signal as input. Three modes can be identified, video, 2:2 pull-down and 3:2 pull-down. While 3:2 pull-down only occurs in 60 Hz video, 2:2 pull-down occurs in both 50 and 60 Hz video.

- Film detectors have a general structure comprising of three blocks: One or more video memories, a pull-down detector and a temporal pattern detector. The pull-down detectors can be classified by the distinctive approach they use. The three approaches are: zero vector matching, horizontal jagged edge detection and signature generation.

- Two types of temporal pull-down patterns can be discriminated, the 'frame difference pattern' and the 'field difference pattern'. The 'frame difference pattern' can detect 3:2 pull-down robustly, but cannot be used to detect 2:2 pull-down. The 'field difference pattern' can be used to detect 2:2 and 3:2 pull-down, but suffers from the fact that the video signal is interlaced. These patterns are more difficult to generate. Hybrid solutions exist that use both patterns.

- The following problems regarding to film detection were observed in the publications:

  False motion detection When detecting 'field difference patterns', the pull-down detector has to deal with interlacing. This causes false detection of motion, breaking the pattern.

  Video composites Video signals may contain multiple areas that each have different film modes. An area in video mode will not be detected, if the amount of motion in the area is small compared to the amount of motion in the rest of the image. This way, the video part is processed in the wrong mode.

  Video edits Editing video in film can cause a disruption of the temporal motion pattern.

  Several solutions have been proposed to counteract these problems.

- Two fields of application can be identified: de-interlacing and encoding. Generally, for encoding, a zero vector matcher using a frame difference is used. This pull-down detector produces frame difference patterns. This is a logical choice, because in encoding, the repeated field in 3:2 pull-down has not to be encoded. The frame difference is a robust method for detecting this redundant field.

- In the field of de-interlacing, progressively scanned sequences have to be distinguished from other sequences, to prevent faulty de-interlacing of non-progressive video material. Progressive video cannot be detected using a frame difference patterns. Therefore, field difference patterns are required to detect progressively scanned sequences.

- Solutions are emerging in the field of de-interlacing that implement zoning and measures that counteract false detection due to false detail.

In the field of film detection, some trends can be discovered. The problem of 3:2 pull-down detection seems to have been sufficiently resolved, since there is one basic implementation that has not improved over the decade. The solution is sufficient for the area of encoding, where they need to detect duplicate fields in a video stream. In that case, an incorrect detection of the film mode will not result in annoying artifacts, but in a less efficient encoding of data.

For de-interlacing, increasingly complex solutions have been proposed for 2:2 pull-down. This indicates that a satisfactory solution has not yet been found. Incorrect detections in de-interlacing applications cause serious artifacts. Therefore the robustness of the detectors has to be much greater than for encoding.

Appendix B

Keyword Search

The result of the keyword search are:

- Lyon, T.C. and Campbell, J.J., MOTION SEQUENCE PATTERN DETECTOR FOR VIDEO, Assignee: Yves C. Faroudja, Los Altos Hills, Calif. US, 1 januari 1991, United States Patent Office, US 4,982,280
- Wenckenbrock, J.H. and Strolle, C.H., Motion detection in television signals, Aspplicant: Samsung Electronics Co Ltd, Republic of Korea, October 10 1993, UK Patent Office, UK Patent Application GB 2 258 580 A
- Casavant, S.D., Hurst, R.N., Perlman, S.S., Isnardi, M.A. and Aschwanden F., VIDEO/FILM-MODE (3:2 PULLDOWN) DETECTOR USING PATTERNS OF TWO-FIELD DIFFERENCES, Assignee: RCA Thomson Licensing Corporation, Princeton, N.J., May 31, 1994, United States Patent Office, US 5,317,398
- Correa, C. and Schweer, R., METHOD AND DEVICE FOR FILM-MODE DETECTION, Assignee: Deutsche Thomson-Brandt GmbH, Villengen-Schwenningen, Germany, Nov. 15, 1994, United States Patent Office, US 5,365,273
- Martin, A., METHOD AND DEVICE FOR FILM-MODE DETECTION, Assignee: Thomson Consumer Electronics, Inc., Indianapolis, Ind., Apr. 11, 1995, United States Patent Office, US 5,406,333
- Martin, A. and Smith, M, METHOD AND DEVICE FOR FILM-MODE DETECTION AND FIELD ELIMINATION, Assignee: Thomson Consumer Electronics, Inc., Indianapolis, Ind, US, Sep. 19. 1995, United States Patent Office, US 5,452,011
- Jun Yonemitsu, Teruhiko Suzuki and Yoichi Yagasaki, APPARATUS FOR CODING AND DECODING A DIGITAL VIDEO SIGNAL DERIVED FROM A MOTION PICTURE FILM SOURCE, Assignee: Sony Corporation, Japan, Oct. 24, 1995, United States Patent Office, US 5,461,420
- Bock, A.M. Removal of redundant fields in standards conversion, Applicant: Digi-Media Vision Limited, London, UK, Oct. 15, 1997, UK Patent Office, UK Patent Application GB 2 312 116 A
- Christopher, T.J. and Correa, C., METHOD AND APPARATUS FOR IDENTIFYING VIDEO FIELDS PRODUCED BY FILM SOURCES, Assignee: Thomson Consumer Electronics, Inc., Indianapolis, Ind., Nov. 18, 1997, United States Patent Office, US 5,689,301
- Yong Sun Lee, Jin Hwan Lee, Yo Sung Ho and Joo Hong Jeon, FILM MODE VIDEO SEQUENCE DETECTOR, Assignee: Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, Mar. 31, 1998, United States Patent Office, US 5,734,420
- Correa, C. and Schweer, R., Verfahren un Vorrichtung zur Film-Mode-Detection, Patentinhalber: DEUTSCHE THOMSON-BRANDT GMBH, Villingen-Schwenningen (DE), Jul. 1, 1998, European Patent Office, EP 0 567 072 B1

- Swartz, P.D., FILM SOURCE VIDEO DETECTION, Faroudja Laboratories, Inc, Sunnyvale, CA, US, Apr. 22, 1999, World Intellectual Property Organization, International Bureau, WO 99/20040
- Yagasaki Yoichi and Suzuki Teruhiko, Coding and decoding digital video signals having duplicate pictures and frames with fields originating from different film source frames, Assingee: Sony Corporation, Tokyo, JP, Eorpean Patent Office, EP 0 588 668 B1
- Coombs, G.R., Derek, A. and Morris, O.J., Identifying film frames in a video sequence, Assignee: Philips Electronics UK Limited, Croydon, GB, Dec. 22, 1999, European Patent Office, EP 0 612 187 B1
- Swartz, P.D., FILM SOURCE VIDEO DETECTION, Assignee: Faroudja Laboratories, Inc., Sunnyvale, Calif. US, Jan. 11, 2000, United States Patent Office, US 6,014,182
- Christopher, T.J. and Correa, C., Method and apparatus for identifying video fields from film sources employing 2-2 and 3-2 pull down sequences, Assignee: Thomson Consumer Electronics, Inc. Indianapolis, In, US, Jan. 26, 2000, European Patent Office, EP 0 720 366 B1
- Christopher, T.J. and Correa, C., Method and apparatus for identifying video fields produced by film sources, Assignee: Thomson Consumer Electronics, Inc, Indianapolis, In, US, Feb. 9, 2000, European Patent Office, EP 0 720 367 B1
- Gerets, P. METHOD AND APPARATUS FOR VIDEO SIGNAL PROCESSING, Applicant: BARCO N.V. Poperinge, BE, Mar. 23, 2000, World Intellectual Property Organization, International Bureau, WO 00/16561
- Swan, P.L., SYSTEM AND METHOD FOR RECONSTRUCTING NONINTERLACED CAPTURED CONTENT FOR DISPLAY ON A PROGRESSIVE SCREEN, Assignee: ATI Technologies, inc., Thornhill, Canada, Apr, 25, 2000, United States Patent Office, US 6,055,018
- Hui Yau Wei Lucas, PROGRESSIVE/INTERLACE AND REDUNDANT FIELD DETECTION FOR ENCODER, Applicant: STMICRO-ELECTRONICS ASIA PACIFIC PTE LTD, Singapore, Jun. 8, 2000, World Intellectual Property Organization, International Bureau, WO 00/33579
- Faroudja, Y.C., Swartz, P.D., Campbell, J.J., HIGH-DEFINITION TELEVISION SIGNAL PROCESSING FOR TRANSMITTING AND RECEIVING A TELEVISION SIGNAL IN A MANNER COMPATIBLE WITH THE PRESENT SYSTEM, Assignee: Faroudja Laboratories, Inc., Sunnyvale, Calif, US, Aug. 22, 2000, United States Patent Office, US 6,108,041
- Swartz, P.D., FILM SOURCE VIDEO DETECTION, Assignee: Faroudja Laboratories, Inc., Sunnyvale, CA, US, Mar, 13, 2001, United States Patent Office, US 6,201,577 B1
- Il Taek Lim, FILM MODE DETECTION METHOD USING PERIODIC PATTERN OF VIDEO SEQUENCE, Assignee: LG Electronics Inc., Jun. 7, 2001, United States Patent Office, United States Patent Application US 2001/0002853 A1
- Del Corso, S., FILM SOURCE VIDEO DETECTION AND ENCODING, Applicant: Koninklijke Philips Electronics N.V., Eindhoven, NL, Sep. 7. 2001, World Intellectual Property Organization, International Bureau, WO 01/65857 A1