MPEG-2 Compliant Trick Play Over a Digital Interface

Onno Eerenberg, Peter H.N. de With, Joep P. van Gassel, Declan P. Kelly

Abstract — Current analog Video Cassette Recorders (VCR) support trick-play modes such as fast search and slow motion respectively. Current analog and digital tape based video recorders support trick play mode such as fast search and slow motion. It is desirable that non-tape based systems have at least the same functionality as the already existing systems. This paper presents an experimental trick-play technique that allows visual search on a remotely locate disk-based storage system connected to a digital TV via a digital interface (see Fig. 1), that resembles the performance of tape-based recorders. Our concept is based avoiding intermediate transcoding of the video signal, but instead re-using the recorded MPEG-2 encoded video. This allows generation of fast search modes in forward and reverse direction, as well as slow-motion in forward direction. A special interlace processing step is introduced, to overcome annoying motion judder when interlaced pictures are re-displayed twice or more. Our experimental results show that the proposed reduced refresh-rate mechanism indeed shows better visual performance than a conventional system without additional processing and it copes with various real-time constraints such as bandwidth, disk seek-rate and CPU load.1

Index Terms — Fast-search, interface-kill, MPEG-2, reduced refresh-rate, slow-motion, trick-play

1 Onno Eerenberg is with Philips Research, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands. (e-mail: onno.eerenberg@Philips.com).
Peter H. N. de With is with LogicaCMG / Eindhoven University of Technology, Den Dolech 2, 5600 MB, The Netherlands (e-mail: p.h.n.de.with@tue.nl).
Joep P. van Gassel is with Philips Research, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands. (e-mail: joep.van.gassel@Philips.com).
Declan P. Kelly is with Philips Research, 38F Tower 1 Kerry Everbright City, 218 Tian Mu Xi Rd. Shanghai, P.R.China 200070. (e-mail: declan.kelly@Philips.com).

I. INTRODUCTION

In the past decade, digital video has become available for consumer applications [1], [2], [3], [4]. A transition from analog to digital video recorders has been made via tape-based systems, which are now increasingly based on non-tape storage media, such as hard disk (HDD) and optical-disk (DVD). Due to the continuous growth in storage capacity, the need for navigation such as fast search is augmenting. State-of-the-art digital storage devices are increasingly found in digital interconnected systems. Interconnectivity of digital storage devices constrains the navigation signal processing. This paper discusses a technique to offer trick-play in such systems, allowing transmission over a digital interface. The technical solution to provide trick-play functionality can be split into two situations, signal processing during record or at play backtime. Depending on the technical solution, the trick-play quality varies for tape and disk-based systems, due to the nature of data storage [5], [6]. For tape-based digital storage systems, such as DV [1] or [2], signal processing during record on the recorded video information is required to enable trick-play at play back. For disk-based storage systems, different strategies can be applied. A popular method is the use of the so-called Characteristic Point Information (CPI). CPI can be seen as a Look-Up-Table (LUT), used to store characteristics of the recorded program. The LUT provides e.g. the physical storage position on disk of the normal-play entry points. The normal-play entry points are defined by the start of intra-encoded pictures at the beginning of an MPEG-2 Group-Of-Pictures (GOP). This paper presents an algorithm to implement fast search in forward and reverse direction and slow-motion in forward direction. Our concept is based avoiding intermediate transcoding of the video signal, but instead re-using the recorded MPEG-2 encoded video [8], [9]. The new derived trick-play video sequence is MPEG-2 compliant [10] and can be transmitted across a digital interface [7]. Furthermore, an efficient algorithm is presented for elimination of motion judder, which occurs when an interlaced image is displayed more than once in succession. The attractiveness of both algorithms is that they are applied in the MPEG-2 compressed domain, avoiding any MPEG-2 decoding and encoding resulting in a low implementation complexity.

The paper is divided as follows. Section II explains the basic understanding of video trick-play and addresses fast search and slow motion trick-play. Section III presents the concept to obtain fast search trick-play based on re-use of the normal play MPEG-2 encoded pictures. Section IV presents the "interface kill", a signal processing step that eliminates motion judder, caused by repetitive display of a picture from an interlaced video sequence. This technology enables the

Fig. 1. A possible digital recorder set-up

Authorized licensed use limited to: Eindhoven University of Technology. Downloaded on September 2, 2009 at 06:21 from IEEE Xplore. Restrictions apply.
trick-play is the play back of audiovisual information at speeds that are different from the recording speed. Let $P_s$ be the play-back speed, then the normal play-back situation is defined for $P_s=1$. Two play-back situations can be distinguished with respect to normal play. Fast play-back is defined as $P_s > 1$, whereas slow motion occurs for $P_s < 1$. Both play-back modes can either be in the forward or reverse direction. In Fig. 2, fast search is obtained by equidistant selection of video frames, where the distance is directly related to the applied search speed. Let us know briefly discuss the difference between fast search and slow motion.

A. Fast search trick-play

Fast search video trick-play is obtained by the aforementioned equidistant sub-sampling of video frames. If $P_s = 4$, a speed-up factor of four is required and thus resulting in a sub-sampling of a factor four. Play back of the new derived video sequence, either forward or reverse, causes the viewer to experience a fast play back speed.

B. Slow motion trick-play

Slow motion video trick-play is obtained by repetitive display of the individual normal play pictures in sequential order. The number of display periods is equal to the reciprocal of the required slow motion speed. Fig. 3 visualizes the temporal display repetition process for a slow motion speed of $P_s = 1/3$.

III. Fast search trick-play on MPEG-2 compressed video

MPEG-2 compliant trick-play across a digital interface can be achieved in two ways. The first approach involves MPEG-2 decoding and encoding (transcoding) of the selected normal play video information. This is an expensive solution, requiring considerable system resources, in particular for fast search trick-play. The second concept is based on re-using of normal play MPEG-2 compressed video information. A starting point is re-use of intra-encoded pictures occurring at the start of an MPEG-2 Group Of Pictures (GOP). As a result, the minimum speed-up factor is equal to the GOP length $N$, provided that a fixed GOP length has been used for encoding the normal play video sequence. Fig. 4 visualizes the temporal sub-sampling process of an MPEG-2 compressed video sequence for a speed-up factor of 12. Note that the subscript used in Fig. 4 indicates a linear index and not the temporal reference as used in the MPEG-2 video compression. The re-use of intra-encoded pictures avoids the needs for MPEG-2 decoding and encoding, but is less attractive than it seems, because the generated trick-play video sequence may not be MPEG-2 compliant. This is caused by the fact that the intra-encoded picture bit-cost can be so high that the maximum allowable bit-rate is exceeded, due to the consecutive transmission of intra-coded information. This problem is eliminated when the transmission of the intra-encoded picture is spread out over more than one display period. However, this solution introduces a video decoder buffer underflow (VBV-buffer violation), because the picture data does not arrived on time. This is due to the fact that the amount of pictures transmitted is less than the frame rate of the trick-play video sequence.
Fast search trick-play with reduced refresh-rate as described above, uses repetition pictures to repeat the last displayed picture. For the situation that the normal-play video sequence has a non-interlaced format (progressive scan), Fig. 9 indicates the temporal reference as used in the MPEG-2 video compression. For the situation that the intra-coded pictures are not skipped, insertion of repetition pictures is a method to generate trick-play sequences with speed-up factors lower than the GOP length $N$. A disadvantage of this method is that the motion is not preserved and further quantized by the repetition process. Fig. 7 presents a snapshot of 240 ms fast search trick-play video with a speed-up factor of 12 and a refresh-rate of 8.33 Hz. At the left column, the intra-encoded pictures selected from the normal play sequence are visualized. The right column shows the generated video trick-play sequence with a speed-up factor equal to 12 and a refresh-rate equal to 8.33 Hz. From this snapshot the consequence of skipping intra-coded pictures becomes clear. Fig. 8 depicts a snapshot of 240 ms fast search trick-play video with a speed-up factor of 4. Each intra-encoded picture selected from the normal play sequence is displayed at the left column. The derived video trick-play sequence is shown at the right column. Each normal play intra-coded picture is displayed three times. The result is a video trick-play sequence with a speed-up factor equal to 4 and a refresh-rate equal to 8.33 Hz.

IV. INTERLACE KILL

Fast search trick-play with reduced refresh-rate described above uses repetition pictures to repeat the last displayed picture. For the situation that the normal-play video sequence has a non-interlaced format (progressive scan), Fig. 9 indicates the temporal dependencies of a predictive encoded repetition picture. Since the registration time of both fields are equal, both fields can be re-displayed via a predicted repetition picture without causing any visual artifacts.
Fig. 7. Snapshot of 240 ms consecutive trick-play video. The pictures are shown in sequential order from top to bottom. (a) This column shows the normal play intra-coded pictures selected from a normal play sequence with a GOP length of 12 resulting in a trick-play video sequence with a speed-up factor of 12 and a refresh-rate of 25 Hz, with the picture transmission time ≤ one display period. (b) This column indicates a trick-play video sequence with speed-up factor equal to 12 based on re-used normal play intra-coded pictures, where each picture requires a transmission time of three display periods resulting in a refresh-rate of 8.33 Hz.

Fig. 8. Snapshot of 240 ms consecutive trick-play video, with the same conditions as in Fig. 7 (a). (a) Same situation as in Fig. 7 (a). (b) This column indicates a trick-play video sequence with speed-up factor equal to four based on re-used normal play intra-coded pictures, where each picture is displayed three times resulting in a refresh-rate of 8.33 Hz.
A problematic situation occurs for interlaced video sequences. If there is any motion in the elapsed time between the registration of the two fields that form the full picture, motion judder occurs. This situation arises, when a predictive encoded picture, as depicted in Fig. 9, is used to repeat an interlaced picture. For the viewer this motion-judder is annoying. A concept to solve this problem, is to eliminate one field and replace this field by the remaining field. Basically, this field elimination, leading to "interlace-kill", can either be done at the recorder side or in the MPEG-2 decoder.

When considering the worse case required signal processing, full MPEG-2 decoding of the intra-coded picture down PCM level, removal of the unwanted field, which depends on the trick-play speed direction, followed by a full MPEG-2 intra encoding step, it is favorable to do this at the receiver side using a predictive coded picture performing the interlace-kill operation. When interlace-kill is applied at the MPEG-2 decoder, it has the lowest complexity. Note that the most important element for applying the interlace-kill depends on the transmission time of the next intra-coded picture. This picture is selected from the normal-play video sequence according to the trick-play speed-up factor. If the bit-cost of this picture exceeds the maximum bit-cost which can be derived from the used MPEG-2 Level, the transmission time is larger than the reciprocal value of the actual frame rate as specified by the MPEG-2 sequence header.

Fig. 10 portrays that interlace-kill is obtained by deleting a field from the original picture, depending on the trick-play speed direction. For the field elimination, an MPEG-2 P- or B-type picture can be used. The difference is that a P-type picture results in a modification of the MPEG-2 decoder anchor buffer contents, whereas a B-type picture does not influence the anchor buffer contents, despite the use of interlace-kill.

Fig. 11 indicates two fields that form one interlace picture.
V. SLOW MOTION ON MPEG-2 COMPRESSED VIDEO

Slow motion is obtained by means of repetitive display of each individual normal play picture, as was explained in Section II.

For interlaced video sequences, special attention should be paid to the original content of the anchor buffers for proper decoding during slow motion. To avoid modification of the anchor buffer contents, only MPEG-2 B-type repetition pictures applying interlace-kill can be used to repeat normal play I- and P-type pictures. The repetition of normal play interlaced B-pictures will give complications, which will be discussed later in this section.

For progressive video, anchor pictures can be repeated using either MPEG-2 P- or B-repetition pictures. Display repetition of normal play MPEG-2 B-pictures, is achieved via re-transmission of that particular picture. Fig. 13 visualizes the generation of slow motion forward, with aid of B-type repetition pictures to repeat the normal play anchor pictures. Hereby, the normal play video sequence is assumed to be progressive scan.

Let us now further analyze the repetition process in terms of the MPEG-2 GOP and profile parameters. For the situation of interlaced video, normal play MPEG-2 compressed B-pictures are basically to be transmitted only once and not repeated, to avoid motion judder. In order to make the normal play B-type pictures suitable for multiple transmission, transcoding is required whereby one field is removed and replaced by the content of the remaining field. Transcoding of normal play B-type pictures is not further considered here, because of the involved partial decoding and encoding of individual fields. The prevention of the above-mentioned motion judder, lead to a trick-play speed error during slow motion, caused by the fact that not every picture is displayed an equal amount of display periods. This situation can be avoided by displaying the anchor pictures more often than required, when compared to a straightforward repetition process. This will become clear from the following analysis.

The GOP length \( N \) is an integer multiple of the P-distance \( M \). The reciprocal of the trick-play slow motion speed \( P_s \) is limited to an integer value. The number of display periods \( D_p \), a normal play picture is displayed during slow motion is specified by

\[
D_p = \frac{1}{P_s},
\]

where \( D_p \) should also be an integer multiple of \( M \). The deviation of the display rate from a normal slow motion repetition process is expressed as a so-called display error \( D_e \). This parameter \( D_e \) indicates the display error per normal play GOP and thus relates to the total number of display periods obtained by a normal play GOP in slow motion \( N \cdot D_p \), so that

\[
D_e = N(D_p \cdot (1-\frac{1}{M}) - (1-\frac{1}{M})).
\]

From (2), it becomes clear that for normal play video sequences, encoded at MPEG-2 simple profile having \( M=1 \), to have zero display error \( D_e \). For MPEG-2 Main Profile with \( M>1 \), the display error \( D_e > 0 \). The speed-error can be compensated via extra display repetitions of the normal play anchor pictures. The number of anchor pictures per GOP is \( N/M \). The extra number of anchor repetitions \( A_e \) is therefore the display error from (2) divided by this fraction, resulting in
\[ A_r = D_p (M - 1) - (M - 1). \]  

For a normal play video sequence with GOP length \( N=12 \) and various values of \( M \), Table I, shows the display error and extra anchor repetitions for a slow motion speed of \( 1/3 \).

### VI. SYSTEM IMPLEMENTATION

#### TABLE I

<table>
<thead>
<tr>
<th>( N )</th>
<th>( M )</th>
<th>( N/M )</th>
<th>( D_r )</th>
<th>( A_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>4</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>3</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>2</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>

We assume a Personal Video Recorder (PVR), based on a Hard Disk Drive (HDD) capable of storing an MPEG-2 transport stream. This system is equipped with Characteristic Point Information (CPI) generation and trick-play functionality as described in the previous sections. Fig. 14 portrays a basic system overview. At the left side of the diagram the compressed MPEG-2 transport stream enters the system. This signal can be viewed real-time, via signal path (a), or in time-shift mode via signal path (b). Play back of a time-shifted program subjected to trick-play, where the normal play signal is processed by the processing blocks in signal path (c) to generate a trick-play signal. At the output of Fig. 14 (at the right side), the transport stream leaves the system and is supplied to an external MPEG-2 decoder. During trick-play, the control block retrieves information from the HDD corresponding to the chosen trick-play mode. During fast search, intra-coded pictures are retrieved from the HDD, based on information stored in CPI. The retrieved information is a snapshot of the multiplexed normal play program, containing the intra-coded picture and other program information e.g. audio and private data. This signal is demultiplexed by the transport stream demultiplexer 'Demux', resulting in a video elementary stream intra-coded picture. This picture is processed by the GOP compositor, 'GOP', which inserts if necessary, interlace-kill repetition pictures and adds the Packetized Elementary Stream (PES) header. The generated trick-play GOP is finally sent to the transport stream multiplexer 'Mux'. This block also adds a Program Association Table (PAT), Program Map Table (PMT) and the Program Clock Reference (PCR), the system time base.

A software-based PVR implementation supporting fast search trick-play was implemented on a media processor\(^2\) DSP. To analyze the DSP CPU load, a 6 Mb/s MPEG-2 transport stream was used as an input. This stream contained audio and video with a GOP length \( N=12 \). Two use cases were explored. The first use case was based on fast search trick-play, with a speed-up factor equal to the normal play GOP length and a picture refresh-rate equal to the frame rate. The second use case was based on fast search trick-play with a speed-up factor equal to 4. In Europe, the refresh-rate for the first use case is equal to 25 Hz and 8.33 Hz for the second use case. Table II indicate the DSP CPU load for both use cases. In Table II, the DSP CPU load is expressed in the required clock frequency (MHz), which also includes the parallelization factor in the DSP. Row one indicates the situation where every normal play I-picture is retrieved from disk has used for fast search, resulting in a refresh-rate of 25 Hz. The DSP CPU load measured indicates the processing for data retrieval, control, demultiplexing, GOP composition and multiplexing. In row two, again each I-picture is recovered, but due to the lower speed-up factor, the GOP compositor inserts two P-repetition pictures resulting in a lower refresh-rate emulating a fast search speed-up factor of four. This reduces the amount of retrieved I-pictures per second with a factor of three, thereby requiring less processing (approximately a factor three).

Summarizing, lowering the refresh-rate can be used to significantly reduce the DSP CPU load. We have found that a reduced refresh-rate is a concept that can be used to address various problems. It provides a proper solution to overcome the HDD seek-rate constraint and it also reduces the load on the system bandwidth. A reduced refresh-rate can also be used to accommodate the visual perception experienced by the viewer during high-speed fast search. Due little or no temporal correlation, the visual information during high-speed trick-play

---

\(^2\) Commercially available as part of a set-top box chip Philips PNX8525.
can hardly be followed by the viewer. Lowering the refresh-rate causes the viewer to better interpret the video information. It should be noted that the proposed trick-play signal processing takes fully place in the MPEG-2 compressed domain and is based on full re-use of the normal play MPEG-2 compressed video pictures.

This concept has a low complexity and therefore provides an efficient implementation.

VII. RESULTS AND CONCLUSIONS

We have presented, a technique for generation MPEG-2 compliant trick-play video sequences, for both fast forward / reverse and slow-motion forward. The derived trick-play sequences fully re-use normal play MPEG-2 encoded video. For the fast search modes, transmission time is created using repetition predictive encoded pictures, allowing full re-use of normal play intra-encoded pictures. For the situation that the video frames are interlaced and a transmission time larger than one display period is required, “interlace-kill” is applied, replacing one field of that frame by the other field, thereby preventing motion judder during display.

For slow motion forward trick-play based on Simple Profile (SP) MPEG-2 coding, simple picture repetition based on B-type coded pictures are used. For Main Profile (MP) MPEG-2 coded video, simple picture repetition lead to motion judder. This can be avoided but results in a trick-play speed-error. By more frequently displaying of the anchor pictures, this speed-error can be fully compensated.

The proposed trick-play signal processing takes place in the MPEG-2 compressed domain and is based on fully reusing of the normal play MPEG-2 compressed video. The concepts can be efficiently implemented with low-complexity signal processing. Since transcoding of received pictures is not required, the signal processing is limited to byte-level parsing of the video elementary stream and picture header interpretation. Additionally, we have found that further system optimizations are possible. For example, experiments have indicated that reducing the refresh-rate is a suitable technique for CPU-load reduction. Reducing the refresh-rate has also a positive influence on the memory-bandwidth usage and storage medium seek-rate. Finally, the reduced refresh-rate gives some perceived quality improvement, because it allows the viewer to interpret the video content for a longer period, which is especially attractive for fast search trick-play with high speed-up factors.

REFERENCES
[8] ISO/IEC 13818-1: "Information Technology - Generic Coding of Moving Pictures and Associated Audio Recommendation H.222.0 (systems)".
[10] ISO/IEC 13818-4: "Information Technology - Generic Coding of Moving Pictures and Associated Audio Recommendation H.222.0 (conformance)".

Onno Eerenberg was born in Zwolle, the Netherlands, in 1966. He graduated from the Polytechnical College in Amsterdam in 1992. He joined Philips Research Laboratories Eindhoven, The Netherlands where he worked in the Magnetic Recording Systems department on digital video and data recording systems. He was involved in several European research projects in this area and was involved in the implementation of e.g. video compression systems. He received a MSc degree in engineering product design in 1998 from the University of Wolverhampton, UK. He is currently working for Philips Research where he is involved in the development of portable DVB receivers and their implementation. He holds several US patents and patent applications in the field of recording.

Peter. H.N de With graduated in electrical engineering from the University of Technology in Eindhoven. In 1992, he received his Ph.D. degree from the University of Technology Delft, The Netherlands, for his work on video bit-rate reduction for recording applications. He joined Philips Research Labs Eindhoven in 1984, where he became a member of the Magnetic Recording Systems Department. From 1985 to 1993, he was involved in several European projects on SDTV and HDTV recording. In this period, he contributed as a principal coding expert to the DV standardization for digital camcording. In 1994, he became a member of the TV Systems group at Philips Research Eindhoven, where he was heading the chair on Digital Circuitry and Simulation with the emphasis on video systems. Since 2000, he is
with CMG Eindhoven (now LogicaCMG) as a principal consultant and he is professor at the University of Technology Eindhoven, at the faculty of Electrical Engineering. He has written and co-authored numerous papers on video coding, architectures and their realization. Regularly, he is a teacher of the Philips Technical Training Centre and for other post-academic courses. In 1995 and 2000, he co-authored papers that received the IEEE CES Transactions Paper Award, and in 2004, the VCIP Best Paper Award. In 1996, he obtained a company Invention Award. In 1997, Philips received the ITVA Award for its contributions to the DV standard. Mr. de With is a senior member of the IEEE, program committee member of the IEEE CES and ICIP, chairman of the Benelux community for Information and Communication Theory, co-editor of the historical book of this community, former scientific board member of CMG, scientific advisor of the Dutch Imaging school ASCII, IEEE ISCE and board member of various working groups.

Declan P Kelly received a BSc in Mathematics and Computer Science from University College Dublin in 1992 and an MSc in Computer Science from Trinity College Dublin in 1996. From 1992 until 1997 he worked for Silicon & Software Systems in Ireland as a software engineer and project leader on a number of projects related to infrastructure for the GSM mobile network. During this time he was also project leader of a CMM based Software Process Improvement initiative. In 1997 he joined Philips Research in The Netherlands and worked on video recording and publishing on optical discs, including Blu-ray Disc standardization. Since 2004 he is based in Philips Research Shanghai and is responsible for a number of storage related projects.

Joep van Gassel received a Master of Science degree in Electrical Engineering from the Technical University of Eindhoven in 1999. After graduating he started working for Philips Research on the application format of the Blu-ray Disc standard. The work particularly focused on MPEG processing for trick play over a digital interface. In 2001 he switched to mobile storage systems and application research. Topics include the system architecture aspects of connectivity, power management, file systems and database technology in the mobile domain.