System design of advanced video navigation reinforced with audible sound in personal video recording

DOI:
10.1109/ICCE.2008.4587879

Document status and date:
Published: 01/01/2008

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

1.3-5

System Design of Advanced Video Navigation Reinforced with Audible Sound in Personal Video Recording

O. Eerenberg[1], R.M. Aarts[2] and P.H.N. de With[3]
[1]NXP Semiconductors Research, Eindhoven, The Netherlands,
[2]Philips Research / University of Technology, Eindhoven, The Netherlands,
[3] LogicaCMG / University of Technology Eindhoven, The Netherlands.

***Abstract*—This paper presents an advanced video navigation concept based on a Picture in Picture (PiP) allowing combined rendering of fragments of normal-play video and fast search video simultaneously. Loss of audio during conventional fast search trick-play is eliminated, by adding the sound of the normal play fragment. The normal-play fragments provide detailed information, whereas the fast-search signal presents a course overview. The audio associated to the normal-play fragments enhances the navigation process, allowing the viewer to concentrate on either the normal play or the fast search window. Efficient signal processing in the MPEG-2 domain permits re-use of the standard MPEG-2 decoding path for decoding the normal play audiovisual fragments. Applying a scalable MPEG-2 decoder for decoding the fast search signal enables a low-cost implementation for Personal Video Recording (PVR).**

I. INTRODUCTION AND SYSTEM REQUIREMENTS

Personal Video Recorders (PVR) with optical disks [1] or Hard Disk Drive (HDD) storage media are equipped with conventional trick-play modes [2]. Drawback of the conventional fast search trick-play is that for high speed-up factors, video is uncorrelated and therefore difficult to interpret and audio signals are lost for presentation. Due to the fast random access of the stored information on disks [3], new non-conventional navigation methods can be implemented. We have investigated a double-window video navigation technique based on a Picture in Picture (PiP) screen, presenting normal-play video fragments in the main window and a trick-mode play such as fast forward or fast reverse video, in the PiP window. The normal-play video fragments are enhanced with the corresponding audio to strengthen the content traceability for the viewer. The rational for the audiovisual fragments is based on the fact that human perception employs both visual and auditory cues, thereby making the scene more vivid. The main window presents normal-play video fragments providing detailed information to the viewer, whereas the PiP window presents fast search video providing an outline of the normal-play video sequence.

II. SYSTEM ASPECTS OF DOUBLE-WINDOW TRICK PLAY

The PVR system contains a hard disk, an MPEG-2 decoder, a video mixer and a control processor. The system is equipped with the conventional navigation methods based on intra-encoded picture selection [2]. A selection of MPEG-2 intra-compressed images and MPEG-2 compressed normal-play audiovisual fragments are retrieved from the storage medium. The normal-play video fragments are MPEG-2 decoded and mixed with a downscaled MPEG-2 decoded fast-search video signal. The resulting PiP is depicted in Fig. 1. The navigation

Fig. 1. The main window shows normal-play fragments with the audio also being rendered corresponding to this window, whereas the PiP window shows video corresponding to the fast search.

method has various system aspects that can be adapted to enhance the viewing appreciation and navigation or compromising the system aspects to reduce system requirements.

- **Trick-play refresh rate.** The trick-play refresh rate is typically equal to the rendering rate. For trick-play with high speed-up factors, it is advantageous to reduce the refresh-rate to allow an improved interpretation by the viewer.
- **Normal-play video fragment duration.** The normal-play fragment duration depends on the perception of a sound signal. Depending on the nature of the sound, this duration will differ. On top of this, interpretation of video also involves minimum time duration.
- **Video trick-play decoding.** A multi-window system displaying independent MPEG-2 decoded video signals requires parallel MPEG-2 video decoders. Various tradeoffs can be made here such as the use of scalable video decoding [4], which will be further explored in Section III.
- **MPEG-2 video fragment processing.** The normal-play fragments are constructed using multiple Group-Of-Pictures (GOP). At the start and end of such a fragment, decoding artifacts may be visible depending on the used MPEG-2 profile. To avoid visible artifacts the first and last normal-play fragment GOP requires specific processing eliminating the cause of decoding artifacts.
- **MPEG-2 audio fragment processing.** Depending on the audio sampling rate, padding samples may occur in an

1-4244-1459-8/08/$25.00 ©2008 IEEE

audio frame. Concatenation of audio frames having padding samples causes an audio buffer overflow, which can be prevented by applying proper signal processing.

III. DESIGN CONSIDERATIONS

Let us assume a HDD based PVR equipped with double-window trick-play, based on the systems aspects as presented earlier. Fig. 2. depicts a basic system block diagram that fulfils the trick-play design aspects, and offers the basic PVR play-back modes. Starting at the left side of Fig. 2., an MPEG-2 Transport Stream (TS) enters the system. When the TS is recorded, *Characteristic Point Information* (CPI) is derived, which allows tracking of the intra-encoded pictures of the recorded program. This information and related parameters are stored in the *meta-database*. This meta-database is stored and retrieved from the storage medium via its own control. The audiovisual information stream is written or retrieved from the storage medium via the *to-disk* block and *from-disk* processing blocks in the diagram. The *read-list* block is used to determine –via the meta-database- the starting point of the reading process of the individual information streams, for the various playback situations. The storage control block that controls the various processing blocks in the diagram is not indicated.

The *video mixer* operates at the picture refresh rate, which is determined by the television standard as decoded by the MPEG-2 decoder. During trick-play, the video mixer combines the pictures of normal-play video fragments with the pictures of the fast search video trick-play. For the situation that the video mixer is not equipped with a video scaling function, or the MPEG-2 trick-play video decoding needs to be carried out by a control processor, it is advantageous to use a scalable MPEG-2 decoder [4]. Such a choice lowers the calculation load and involved memory footprint. Further reduction of the processor load is obtained by applying a temporal reduction factor, thereby lowering the trick-play refresh rate, resulting in fewer pictures to be decoded by the *trick-play video decoder*. A reduced refresh rate also lowers the hard disk seek-rate, which improves the Mean-Time-Between-Failure (MBTF). The read-list block is responsible for controlling the from-disk processing block, which retrieves the normal-play TS or normal-play TS fragments. In trick-play mode, the read-list block selects TS fragments containing intra-encoded pictures, using the meta-database, that form the fast-search video trick-play signal. When doing so, the selection process takes into account the temporal reduction factor for an appealing presentation. Besides this, the read-list block also retrieves normal-play fragments.

The *audiovisual trick-play processing* block performs an MPEG-2 de-multiplex operation to access the individual MPEG-2 compressed audiovisual access units. The video processing operates on the retrieved video access units by eliminating pictures, using signal processing in the MPEG-2 domain. The signal processing replaces video access units that refer to non-available reference pictures, thereby avoiding decoding artifacts at fragment boundaries. This block also performs the selection of the compressed normal-play audio frames associated to the video fragments, avoiding bit-rates violations. After processing the audiovisual access units, the individual elementary streams are multiplexed into an MPEG-2 TS, which are supplied to the MPEG-2 audiovisual decoder.

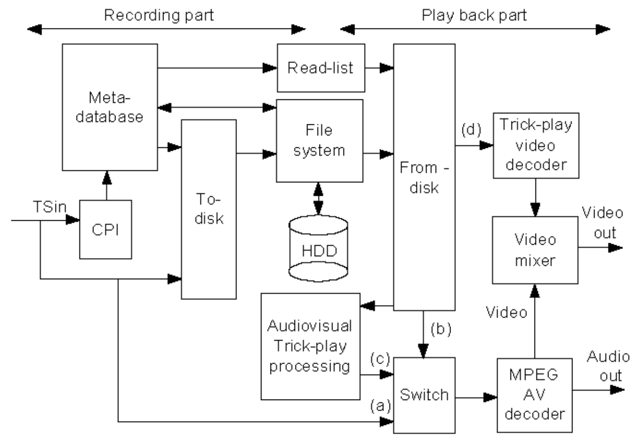

Fig. 2. HDD-based PVR architecture. Switch in position (a) indicates the real-time viewing mode; in case (b), the play-back of delay viewing; and in case (c), the trick-play operation mode for the normal play trick-play fragments, the fast search trick-play is flow in indicated by (d).

IV. CONCLUSIONS

We have presented a new navigation method using a double video window, which combines normal-play fragments and fast search video. The presence of normal-play sound information enhances the proper finding of a video passage by the viewer. The benefit of our proposal is that an outline of the recording is obtained via the fast search trick-play sequence, while the normal-play video fragment refines the perceived impression. Subjective evaluations have indicated that the sound signal guides the viewer in selecting the proper window. Although the system has a double independent video window, the system resources are kept within acceptable boundaries, by applying scalable video decoding techniques for the trick-play video decoder. To lower system resources even further, trick-play with reduced refresh rates can be applied. The elimination of motion-compensated pictures referring to non-available reference pictures and adaptation of the sound signal, allows the usage of a standard MPEG-2 decoder, without the need for interaction to prevent decoding artifacts.

REFERENCES

[1] C. Buma, R. Brondijk and S. Stan, "DVD+RW: 2-Way Compatibility for Video and Data Applications", Proc. of ICCE 2000, pp. 88-89, 2000

[2] O. Eerenberg, P.H.N. de With, J. van Gassel and D.P. Kelly, "MPEG-2 compliant trick play over a digital interface", IEEE Trans. Consumer Electronics, Vol. 51, No.3, pp. 958-966, Aug. 2005.

[3] ISO/IEC 13818-1/2/3: "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information: systems/video/audio".

[4] C. Yingwei, Z. Zhun, L. Tse-Hua, S. Peng, K. van Zon, "Regulated complexity scalable MPEG-2 video decoding for media processors", IEEE Trans. On Circuits And Systems For Video Technology, vol. 12, no. 8, Aug. 2002, pp. 678-687.