A video quality improvement technique for a fast combined watermark embedding method

Hajime Matsunaga¹, Tomoko Sawabe², and Masami Kihara²a)

¹ Graduate School of Science and Engineering, Nihon University,
7–24–1 Narashinodai, Funabashi 274–8501, Japan
² Department of Computer Engineering, College of Science and Technology,
Nihon University, 7–24–1 Narashinodai, Funabashi 274–8501, Japan
a) kihara.masami@nihon-u.ac.jp

Abstract: The use of watermarks to embed user information is expected to deter the unauthorized use of video content in VOD (Video On-Demand) streaming services. This requires that watermark embedding and video encoding for each video must be finished so rapidly as to not create excessive VOD delay between the user’s video request and video download to the user. This paper describes a watermark embedding and encoding process for a fast combined watermark embedding method, and proposes a new encoding technique that minimizes video quality degradation.

Keywords: VOD service, video, watermark embedding, encoding

Classification: Multimedia Systems for Communications

References

[1] H. Yamada, T. Sugai, H. Nishikawa, H. Ohmata, K. Majima, M. Kurozumi, and T. Inoue, “Quick insertion and extraction of watermarks using GPU,” ITE Annual Convention 2011, 11-4, 2011. DOI:10.11485/iteac.2011.0.11-4-1
[2] J. Naganuma, “System LSIs for digital broadcasting applications,” NTT Tech. Rev., vol. 5, no. 10, Oct. 2007.
[3] A. Kobayashi, A. Hasegawa, and M. Fukuda, “A practice of copyright protection applying digital watermark technology (Studying applications in still pictures and moving pictures, and exploration of future expectations),” J. Inf. Process. Manage., vol. 60, no. 2, pp. 89–99, May 2017. DOI:10.1241/joikanri.60.89
[4] A. Giladi, “Integrating forensic watermarking into adaptive streaming workflow,” IBC2017 (International Broadcasting Convention), Oct. 2017.
[5] H. Nagasawa and M. Kihara, “Video content individualization based on quasi-digital watermarking,” Proc. 2015 IEICE Society Conference, B-7-37, Sept. 2015.
[6] Z. Wang, “The SSIM index for image quality assessment,” http://www.cns.nyu.edu/~lcv/ssim/.
[7] A. Hore and D. Ziou, “Image quality metrics: PSNR vs. SSIM,” 2010 20th International Conference on Pattern Recognition, pp. 2366–2369, 2010.
1 Introduction

VOD (Video on-demand) services such as dTV, hulu, Netflix and Amazon Video are increasing. These services provide movies, dramas, and TV programs, and interest is growing in 4K UHD (Ultra High Definition) content with the high resolution of $3840 \times 2160$. While the encryption systems such as HDCP (High-bandwidth Digital Content Protection system) used in HDMI (High Definition Multimedia Interface) and the new DRM (Digital Rights Management) mechanism called a CDM (Content Decryption Module) solutions such as Google Widevine, Apple Fair Play and Microsoft PlayReady in video service systems attempt to prevent the unauthorized usage of video contents, the illegal upload and download of video content will not disappear. To supplement existing DRM techniques, we can use watermark technology to invisibly embed user information into the video content. The embedded user information can be extracted from the video content released illegally.

Unfortunately, the real-time processing needed to embed watermarks into content demands high performance processors [1] or specialized custom LSIs [2]. The watermark combination method was developed to solve this problem [3, 4, 5]. The method well supports VOD services by minimizing the watermark embedding time and encoding time.

This paper proposes a fast watermark embedding method based on combining information that is embedded in the IDR (Instantaneous Decoder Refresh) frames of h.264. Moreover, it introduces an encoding technique that minimizes the degradation in video quality achieved by the proposed combined watermark embedding method.

2 Conventional watermark embedding

2.1 Basic watermark embedding method

The basic watermark method directly embeds user-specific information into video frames; the example of “101” is shown in Fig. 1(a-1). The video content is encoded after the user-specific information is embedded (Fig. 1(a-2)). The encoding is executed using the IDR (Instantaneous Decoder Refresh) frames of h.264 or GOP (Group of Picture) frames in MPEG-2. These units are called “closed units” in this paper. While this approach offers robust embedding of user-specific information, its time overhead is a barrier to adoption given the real-time processing requirements of VOD services.

2.2 Watermark combination method

The watermark combination method [3, 4, 5] was developed to shorten the watermark embedding and encoding time, i.e. the barrier described in Section 2.1. The watermark combination method does not directly embed the complete user-specific
information but uses a sequence of basic identifiers. The user information of “101” can be specified by the sequence of the three basic identifiers “1”, “0” and “1” as shown in Fig. 1(b-1). The unit of the sequence as shown in Fig. 1(b-1) is called the combination unit in this paper. Basic identifiers are embedded in the head frames of the closed unit, called ID frames in this paper. In the example given, the combination unit is a sequence of 3 head frames, and the number of basic identifiers in each ID frame is 2, so this example can distinguish up to 8 users.

After a user request is received, the video data to be sent to the user is appropriately formed by simply selecting the corresponding stock sequential frames as shown in Fig. 1(b-1). This watermark embedding process is thus very fast.

3 Proposed combined watermark embedding and encoding method

Fig. 2 shows an example of the proposed method. The original video frames, that are decoded from compressed video data, are shown in Fig. 2(c-1). The ID frames are extracted from the decoded video frames in Fig. 2(c-1), and as many copies as there are basic identifiers are made. The basic identifiers are then embedded into the copied ID frames by conventional watermark software (Fig. 2(c-2)). In the other watermark embedding approach, all video frames are prepared from the decoded video frames, and basic identifiers are embedded to the ID frames as an initial watermark (indicated by “X” in Fig. 2(c-3)). The initial watermark embedding
operation is very important in determining the final video quality (the advantage of the initial watermark is described in Section 5).

Each ID frame with its basic identifier (Fig. 2(c-2)) is encoded one by one (Fig. 2(c-4)) to yield the IDR frames. All frames with the initial watermark (Fig. 2(c-3)) are also encoded based on the closed unit (Fig. 2(c-5)). This is performed in a preprocessing step before a user requests a video.

When a user request is received, the sequence of the encoded ID frames with the appropriate basic identifier is decided. The selected encoded ID frames with the basic identifier (“1”, “0” and “1” in Fig. 2(c-6)) and all encoded frames in closed units excluding ID frames are combined which yields the video file to be downloaded.
4 Processing time estimation

Two 4 second 4K video sequences, 30 frames per second and compressed to yield 50 Mb/s, are used to estimate the processing time of the proposed method. Since the video data reconfigured as shown in Fig. 2(c-6) is not held as sequential data in the storage media, random access is performed.

We estimate that the encoded ID frames and closed units excluding ID frames for 1 closed unit (30 frames for 1 second) have the read times of approximately 42 ms and 62.5 ms, respectively, and that the maximum total read time is 144.5 ms including average access time of 20 ms. Since this read time is shorter than the length of 1 closed unit, this implies that the video data assumed here can be smoothly played with an adequate (several seconds) video buffer.

In this estimation we used a HDD (Hard Disk Drive) as the storage medium. Its I/O data-transfer rate is 600 MB/s max, spindle speed is 5400 rpm and cache buffer is 64 MB.

On the other hand, in the conventional method as shown in Fig. 1(a-1), we have to consider that the watermark embedding time and the encoding time for 4K video data processing are around 3 seconds per frame using the conventional watermark software and 2 to 7 seconds for a closed unit (30 frames for 1 second), respectively. HDD read and write times are excluded.

These times mean that, without the proposed method, the user has to wait at least 5 to 10 seconds until the video data is played, even if the video has just one-second of screen time. This waiting time is seen as excessive in normal video services.

The computer used in the estimation and experiment had an Intel Core i7-3770 3.4 GHz processor and DDR3 32 GB memory with Windows 10 OS for the encoding and CentOS Linux 7.1 for watermark embedding. The encoding process fully uses the computer’s 4 cores and 8 threads.

5 Video quality evaluation

While the proposed method can eliminate the watermark embedding time and the encoding time, the reconfiguration demanded by inserting the encoded ID frames degrades the video quality. This section describes the results of a video quality evaluation when the proposed method is applied to 4K video.

The video quality in decoded frames was evaluated by the typical measures of SSIM (Structural Similarity) [6] and PSNR (Peak Signal-to-Noise Ratio) [7]. These measures were calculated by FFmpeg, a library of multimedia data transcoding functions [8].

Fig. 3 shows the results of a video quality evaluation using SSIM and PSNR. Two 4 second 4K video data (30 frames per second, 50 Mb/s) were used in this evaluation. Video-A shows a slow panning shot in a room with plain color walls. Video-B shows a landscape including cherry trees with many cherry blossom petals flying around. Video-A and Video-B were encoded with the h.264 codec with FFmpeg and their encoded data rate was around 50 Mb/s.

The characteristics of three encoding methods are compared in Fig. 3: the method with the initial watermark embedding technique described in Section 3
(blue points in Fig. 3), the method without initial watermark embedding (orange points in Fig. 3) and additionally the method that encodes all frames including all ID frames configuring the closed unit (gray points in Fig. 3). Horizontal axis in Fig. 3 is frame number. Frame numbers 1, 31, 61 and 91 are IDR frames.

![Fig. 3. Results of video quality evaluation](image)

Video data encoded with all frames (gray points) has the highest SSIM and PSNR in both Video-A and Video-B, since all frames including IDR frames are encoded. The method that encodes all frames is the best approach considering video quality, however the prepared ID frames, which equals all frames of the closed unit, generates and stores a large amount of data in the preprocessing stage. That is why we propose the method with the initial watermark; it can minimize the volume of the encoded ID frames including user-specific information that needs to be held in the video data storage media.

In the encoding method without initial watermark embedding, SSIM and PSNR are greatly degraded for about 10 frames from the IDR frame in both Video-A and Video-B. However, in the encoding method with initial watermark embedding, SSIM is improved to around 0.98 and an increase of about 1 to 4 dB in PSNR was attained. The proposed method, described in Section 3, is very effective in improving decoded video quality.

We also carried out a subjective assessment based on 15 participants making dichotomous choices. The test, which compared the method of encoding all frames with the method with the initial watermark, found no significant difference between the resulting videos, though it is obvious that the method without the initial watermark yielded worse quality than the method of encoding all frames. Qual-
pressive assessments on a 4K display also confirmed that the proposed method created no perceptible degradation in video quality.

In comparisons with and without the initial watermark in Video-A, while 73 percent of the participants found the proposed method to be superior to the method without the initial watermark, no significant improvement appeared in the evaluation results for Video-B. While the proposed method is effective, the degree of video quality improvement varies with the video content.

6 Conclusion

This paper proposed a method for embedding watermarks into video data by the technique of embedding initial watermarks in ID frames. The method can minimize the video quality degradation triggered by the video frame reconfiguration that results from the insertion of ID frames in the combined watermark combination method. Tests on two 4K videos showed that the proposed technique is very effective.