Improved Techniques for Dual-Bitstream MPEG Video Streaming with VCR Functionalities

Shih-Yu Huang
Department of Computer Science and Information Engineering, Ming Chuan University
5 Teh-Ming Rd., Gwei Shan District, Taoyuan Country 333, Taiwan

Abstract

In this paper, a new dual-bitstream technique for video cassette recording (VCR) functionality and a transcoding technique of the motion vectors (*mv*s) between dual bitstreams are introduced. In the dual-bitstream technique, a direct reference bitstream is proposed to replace the original reverse bitstream to reduce the required network bandwidth and decoder effort of VCR requests. In the transcoding technique, the *mv*s in the forward bitstream are first reused to generate the *mv*s in the reverse bitstream and the direct reference bitstream. If the obtained error is larger than a preset threshold, the *mv*s are refined step-by-step from the inner to the outer of a refinement window.

Index terms -- video cassette recording (VCR), MPEG video, transcoding, motion vectors.

1. Introduction

Multimedia delivery over the Internet draws more attentions nowadays since the progress in the multimedia compression and broadband networking technologies. These multimedia applications such as video-on-demand (VoD) allow users to access the remote videos over the networks by means of streaming technology. To accelerate the marketability of VoD services, a key issue is to provide VCR functionality to enable users interactively browsing the multimedia content. The set of VCR operations include forward, backward, stop, pause, random-access, fast-forward and fast-backward at minimum.

In 2001, a dual-bitstream technique is proposed in [1] for MPEG video streaming with VCR functionality. Fig. 1 (a) is a demonstrative structure of the dual bitstreams in which the video is coded in I-/P-frames with a group of picture (GOP) size of N (=14) frames. Besides the traditional forward bitstream (FB), a reverse bitstream (RB) is added in the server for the reverse play, where the I-frames in the RB are interleaved between I-frames in the FB. By switching frames between the FB and RB, a novel frame-selection scheme based on a least-cost criterion is proposed to determine the transmitted frames over the network for any speedup factors. Comparing to the traditional single forward bitstream, the dual-stream technique reduces that the required network bandwidth and decoder effort to carry VCR commands. The average number of frames to be decoded is reduced from 7.5 to 2.71 for N=14 by using the dual-bitstream technique.

This work was supported by the National Science Council, R.O.C., under Grant NSC 90-2213-E-130-002.

The dual-bitstream technique suffers two problems. First, it would cause the drift problem because the P-frames in the reverse (forward) bitstream would be approximated by the I-frames in the forward (reverse) bitstream during VCR commands. Although the drift problem could be compensated by additional drift bitstreams, it's too complicated. Second, the required network bandwidth and decoder effort can be further reduced. In this technique, the number of decoding frames for a requested frame is still proportional to the distance to the nearest I-frame in one of the FB and the RB. The worst case occurs for the frames located at the middle of two I-frames. About N/4 frames should be decoded to access the middle P-frames. For the structure shown in Fig. 1 (a), Frame 10 is a demonstrative example where Frame 7, Frame 8, Frame 9 should also be first decoded to access this frame.

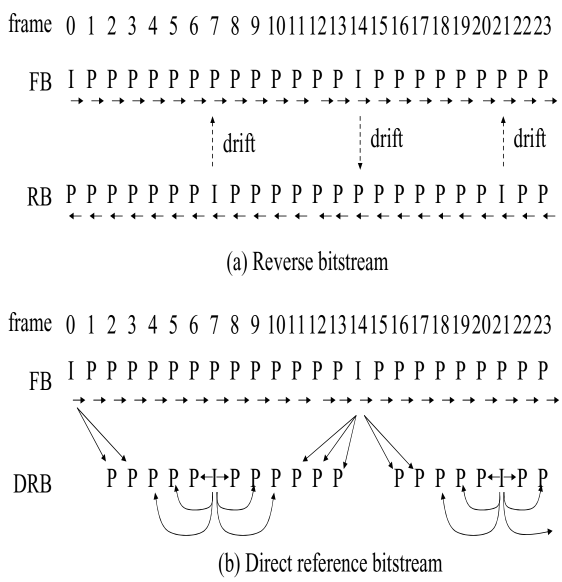


Fig. 1. The structure of the dual-bitstreams.

Based on the concept of dual bitstreams, a direct reference bitstream (DRB) shown in Fig. 1 (b) is introduced to solve the above problems in this paper. The I-frames in the FB are reused in the DRB, i.e, these frames implicitly exist in the DRS. The I-frames in the RB are also added in the DRB. The other frames in the DRB are P-frames directly referenced to their nearest I-frames in the FB or the DRB. Therefore, any P-frames can be directly decoded by the corresponding I-frames. It should be noted that the P-frames next to the I-frames in the FB could also be reused in the DRB. The required space of the DRB is smaller than that of the RB. The saving space is about the sizes of two P-frames for a GOP.

The proposed scheme doesn't suffer the drift problem because the approximation between the I-frames and the P-frames on the dual bitstreams don't exist. The required network bandwidth and decoder effort can be further reduced by the proposed scheme because the number of decoding frames to access any I-frames and P-frames are restricted to be 1 and 2, respectively. The average number of frames to be transmitted is 1.86 for N=14 by using the proposed method. About 30% (= 1 -1.86/2.71) frames could be saved on the average.

In [1], the RB is originally generated by encoding the video frames in the reverse order. Actually, the transcoding techniques [2-3] could be utilized to reduce the required computations especially for the motion estimation taken about 60-80% of total computation in a video encoder. Instead to re-compute the motion vectors (*mv*s) in the RB, a straightforward transcoding method for the above *mv*s is to inverse the *mv*s in the FBs. However, the inverse technique is not suitable for the sequences with large motion contents. A refinement method of *mv*s is presented when the PSNR results generated by the inverse technique cannot be accepted. The above transcoding technique also employed to generate the *mv*s in the proposed DRB. The experimental results show that the proposed technique could greatly reduce the computation and achieve high quality of *mv*s.

The rest of this paper is organized as follows. Section 2 briefly summarizes the original and proposed dual-bitstream techniques. Section 3 gives the transcoding method of motion vectors. Some conclusions are drawn in Section 4.

2. Dual-bitstream Techniques

A. The Reverse Bitstream

The original technique has dual bitstreams: a FB and a RB. The order of the RB is opposite to that of the FB and the I-frames in the RB are interleaved between I-frames in the FB. Therefore, the dual bitstreams have two I-frames in a GOP. If the requested frame is an I-frame in one of the two bitstreams, the frame can be decoded by itself. No extra frames are necessary to be transmitted. If the requested frame is a P-frame, a nearest I-frame either in the FB or the RB is first selected, where the distance is measured by the number of frames from the requested frame to the I-frames in the dual bitstreams. To decode the requested P-frame, the selected I-frame is first decoded. The following decoded frames are dependant on the positions of the selected I-frame and the requested frame. If the requested frame is on the right side of the selected I-frame, all P-frames next to the selected I-frame to this requested frame in the FB should be transmitted and decoded. Otherwise, all P-frames previous to the selected I-frame to this requested frame in the RB should be transmitted and decoded. It should be noted that the P-frames in the reverse (forward) bitstream are approximated by the I-frames in the forward (reverse) bitstream during VCR commands. This will cause some drift.

B. The Direct Reference Bitstream

The proposed technique also has dual bitstreams: a FB and a DRB. Similarly, the dual bitstreams have two I-frames in a GOP. The one is the first frame in the FB and the other is the $(\frac{N}{2})^{th}$ frame in the DRB. The I-frame and the first P-frame in the FB are reused in the DRB. The other frames in the DRB are P-frames direct referenced to their nearest I-frames either in the FB or the DRB, where the number of frames from the requested frame to the I-frames in the dual bitstreams is also employed to measure the distance. Therefore, if the requested frame is an I-frame either in the FB or in the DRB, the server only needs to transmit the frame. If the requested frame is a P-frame, it can be decoded by the referenced I-frame and itself. The approximation between the I-frames and the P-frames on the dual bitstreams doesn't exist. Therefore, the proposed technique wouldn't suffer the drift problem.

Table 1 summarizes the number of the decoding frames in the random-access mode for the above two dual-bitstream techniques with N=14. An average of 2.71 frames and 1.86 frames should be transmitted to decode a requested frame for the dual-bitstream techniques with RB and DRB, respectively. About 30% (= 1 -1.86/2.71) frames could be saved in the proposed technique on the average.

Table 1. The number of decoding frames to access any frames. (RB: reverse bitstream, DRB: direct referenced bitstream)

	F7	F8	F9	F10	F11	F12	F13	F14
RB	1	2	3	4	4	3	2	1
DRB	1	2	2	2	2	2	2	1
	F15	**F16**	**F17**	**F18**	**F19**	**F20**	**F21**	**Aveg**
RB	2	3	4	4	3	2	1	2.6
DRB	2	2	2	2	2	2	1	1.8

3. Motion Vectors Transcoding Techniques for Dual-bitstreams

A. The Reverse Bitstream

The *mv*s in the RB could be obtained by inversing the *mv*s in the FBs [2]. Fig. 2 is a demonstrative example to illustrate the transcoding of the *mv*s. Assuming the *mv* of Block(1) from Frame(n) to Frame(n-1) is *MV* and the majority vector selection rule is employed to assist the decision of the *mv*s from Frame(n-1) to Frame(n). The *mv* of Block(4) between Frame(n-1) and Frame(n) in the RB could be set to *–MV* because the majority of referenced block is located at Block(4) of Frame(n-1) in this example,.

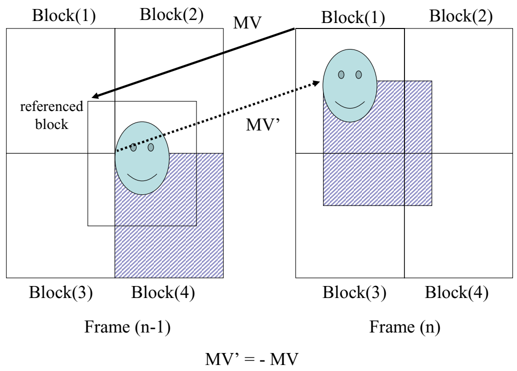


Fig. 2. The inverse *mv* structure for RB.

The inverse technique of *mv*s is suitable for the sequences with small or moderate motion contents. When the motion activities of videos are large, the technique couldn't work very well. New *mv*s for RB need to be computed. To achieve the high quality *mv*s in the RB, the inverse technique is first employed. If the obtained error is smaller than a preset threshold, the algorithm stops. Otherwise, the *mv* is refined within a search range ±1. If the obtained error is still larger than the threshold, the refined range is extended by 1. The steps of refinement are repeated until the obtained error can be accepted or the refined range is equal to the preset search range. The refinement steps are summarized in Fig. 3, where the center of the refinement window indicated as 0 is the location of the inverse *mv*s in the FBs.

-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
7	6	6	6	6	6	6	6	6	6	6	6	6	6	7
7	6	5	5	5	5	5	5	5	5	5	5	5	6	7
7	6	5	4	4	4	4	4	4	4	4	4	5	6	7
7	6	5	4	3	3	3	3	3	3	3	4	5	6	7
7	6	5	4	3	2	2	2	2	2	3	4	5	6	7
7	6	5	4	3	2	1	1	1	2	3	4	5	6	7
7	6	5	4	3	2	1	0	1	2	3	4	5	6	7
7	6	5	4	3	2	1	1	1	2	3	4	5	6	7
7	6	5	4	3	2	2	2	2	2	3	4	5	6	7
7	6	5	4	3	3	3	3	3	3	3	4	5	6	7
7	6	5	4	4	4	4	4	4	4	4	4	5	6	7
7	6	5	5	5	5	5	5	5	5	5	5	5	6	7
7	6	6	6	6	6	6	6	6	6	6	6	6	6	7
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7

Fig. 3. The structure of the refinement steps with a search range ±7.

There is a case that the transcoding technique couldn't be applied. When Frame(n) is an I-frame, the *mv*s from Frame(n) to Frame(n-1) don't exist in the FB. So, the *mv*s from Frame(n-1) to Frame(n) should be recomputed. In this paper, the above refinement algorithm with initial *mv* (0,0) is also employed to recomputed the *mv*s.

We use checking points to evaluate the required computation. One checking point represents the computation taken to check one candidate of the *mv*. For a full search with a search range ±7, 225 checking points are required. For the transcoding technique, one checking point is necessary if the algorithm stops at step 0. For the refinement step *i*, the required checking points are $8\times i=((2i+1)^2-(2i-1)^2)$.

Both the inverse and the proposed methods are implemented to exhibit the performance of the *mv*s transcoding techniques. Full searching of the *mv*s with search range in ±7 in the reverse order of frames is also implemented for comparison. Table 2 summaries the corresponding results of PSNR values and the number of checking points. Although the required checking point of inverse technique is only one, the corresponding PSNR value is bad for the *football* sequence. The PSNR degradation of the proposed method over the full search is very small. The average degradation is 0.2 dB but only about 1% ~ 6% computation is required. The proposed technique could greatly reduce the computation and achieve high quality of *mv*s.

Table 2. The comparisons of various transcoding techniques for RB.

(a) The PSNR values.

	Inverse	Proposed	Full Search
claire	41.14	41.69	41.87
foreman	33.52	34.85	35.01
football	16.90	22.91	23.15

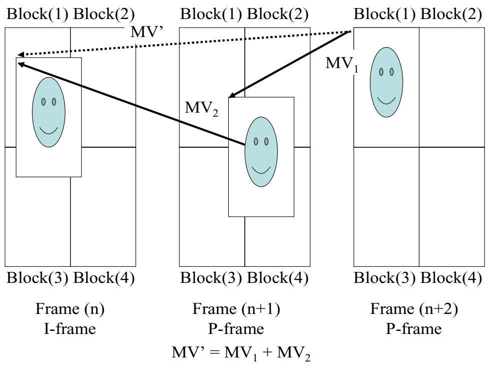
(b) The number of checking points

	Inverse	Proposed	Full Search
claire	1	1.91	225
foreman	1	3.10	225
football	1	12.79	225

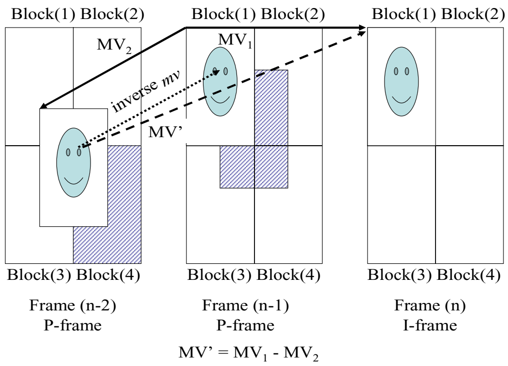
B. The Direct Reference Bitstream

In the DRB, the P-frames could be classified into two categories: the forward P-frames (the referenced I-frames are before to the frames) and the backward P-frames (the referenced I-frames are after to the frames). The *mv*s of the frames belong to the first category could be obtained by the reuse of *mv*s in the FB [3]. Suppose Frame(n) is an I-frame and Frame(n+1) and Frame(n+2) are the forward P-frames referenced to Frame(n). Because the *mv*s from Frame(n+1) to Frame(n) already exist in the FB, they can be reused in the DRB. Fig. 4 (a) is a simple demonstrative example to get the *mv*s from Frame(n+2) to Frame(n). Let MV_1 be the *mv* of Block(1) from Frame(n+2) to Frame(n+1) and the corresponding *mv* from Frame(n+1) to Frame(n) be MV_2

in the FB. Then, the *mv* of Block(1) from Frame(n+2) to Frame(n), *MV'*, in the DRB is the sum of MV_1 and MV_2, i.e., $MV' = MV_1 + MV_2$. Of course, when the obtained error couldn't be accepted, the *MV'* should be refined by the technique mentioned in the above subsection. Based on the same the reuse and refinement techniques, the *mv*s of the other forward P-frames could be generated.



(a). The forward P-frames.



(b) The backward P-frames.

Fig. 4. The reuse *mv*s structure for DRB

Fig. 4 (b) gives the structure to generate the *mv*s for the backward P-frames in the DRB. Similarly, suppose Frame(n) is an I-frame and Frame(n-1) and Frame(n-2) are the backward P-frames referenced to Frame(n). The *mv*s from Frame(n-1) to Frame(n) should be recomputed because the *mv*s from Frame(n) to Frame(n-1) don't exist in the FB. The recomputed *mv*s and the *mv*s from Frame(n-1) to Frame(n-2) in the FB can be reused to generate the *mv*s from Frame(n-2) to Frame(n). Let MV_1 and MV_2 be the recomputed *mv* of Block(1) from Frame(n-1) to Frame(n) and the corresponding *mv* from Frame(n-1) to Frame(n-2) in the FB, respectively. Then, the *mv* of Block(4) from Frame(n-2) to Frame(n), *MV'*, is the sum of MV_1 and the inverse of MV_2, i.e., $MV' = MV_1 - MV_2$, where Block(4) is determined by the majority vector selection rule. Similarly, the *MV'* should be refined when the obtained error is greater than a preset threshold. The *mv*s of the other backward P-frames could also be obtained based on the same the reuse and refinement techniques.

Table 3 gives the results of PSNR values and the number of checking points. The experimental results indicate that the proposed transcoding technique is efficient.

Table 3. The comparisons of various transcoding techniques for DRB, where FP and BP represent the forward P-frames and backward P-frames .

(a) The PSNR values.

		Proposed	Full search
claire	FP	37.04	37.25
	BP	37.47	37.77
foreman	FP	32.36	32.78
	BP	32.45	32.88
football	FP	19.85	19.98
	BP	19.60	20.22

(b) The number of checking points.

		Proposed	Full search
claire	FP	2.41	255
	BP	2.32	255
foreman	FP	4.57	255
	BP	4.43	255
football	FP	14.67	255
	BP	14.28	255

4. Conclusions

In this paper, we proposed two techniques to improve the performance of the dual-bitstream scheme. First, the RB is replaced by the DRB in which the I-frame in the FB are reused in the DRB, a new I-frame is added in the middle of a GOP and all P-frames are directly referenced to their nearest I-frames. Second, a transcoding technique of the *mv*s was proposed for the RB and DRB. The *mv*s of the RB and DRB are initially set to be the inverse and the sum of the corresponding *mv*s in the FB, respectively. A refinement is then applied step-by-step from the inner to the outer of a refinement window until the obtained error can be accepted. The above two techniques greatly improve the performance of the conventional dual-bitstream scheme. Therefore, the proposed techniques are efficient alternatives for MPEG streaming with VCR functionality.

5. References

[1] C. W. Lin, J. Zhou, J. Youn, and M. T. Sun, "MPEG video streaming with VCR functionality," *IEEE Trans. Ciruits Syst. Video Technol.,* vol. 11, no. 3, pp. 415-425, March. 2001.

[2] S. J. Wee, "Reversing motion vector fields," *in Proc. IEEE Int. Conf. Image Processing,* pp. 209–212, Oct. 1998.

[3] T. Shanableh and M. Ghanbari, "Heterogeneous Video Transcoding to Lower Spatio-Temporal Resolutions and Different Encoding Formats, " *IEEE Trans. Multimedia,* vol. 2, no. 2, pp. 101-110, June 2000.