Support Fast Scan Operations with Video Streaming Technology

Chun-Ming Huang\textsuperscript{1}, Kai-Chao Yang\textsuperscript{2}, and Jia-Shung Wang\textsuperscript{3}
Department of Computer Science, National Tsing Hua University
Hsinchu, Taiwan 30043
\textsuperscript{1}cmhuang@cic.org.tw; \textsuperscript{2}kcyang@vc.cs.nthu.edu.tw; \textsuperscript{3}jswang@cs.nthu.edu.tw

Abstract

In this article, we present a novel approach to support fast scan operations for video streaming applications. Our approach is based upon reshaping the ordinary linear encoded GOP structure into a hierarchically encoded binary tree, along with a proximity-based approximation scheme. Our scheme can support forward and backward fast playback operations at any speed-up factor with few bandwidth requirements even for the normal speed playback.

1. Introduction

Rapid advances in multimedia and communication technologies have made content delivery services, such as interactive TV, video-on-demand (VOD), etc. more and more practicable. Apart from normal playback, it is also highly desirable that interactive VCR operations can be provided at affordable cost. Normally the VCR operations include stop-resume, pause-resume, slow motion, jump (random access), fast forward scan (FFS), and fast backward scan (FBS).

To reduce the storage, I/O bandwidth, and translation requirements, most of the VOD systems utilize pre-compressed video streams as the transmitting media. The MPEG video coding standards \cite{1,2} are currently widely used for video compression. The I-P-B frame structure used in MPEG can remove temporal redundancies among video frames very efficiently. But due to the coding dependencies among frames, it is a non-trivial task to support various VCR operations without causing any overhead.

Several approaches have been proposed to support fast scanning operations video streaming. In \cite{3}, fast scan operations are implemented at the client side by using pre-fetched video frames. In \cite{4}, a GOP-skipping-based dynamic transmission scheme was proposed. In \cite{5}, a dual-bitstream approach was proposed that combine an ordinary encoded and a reverse encoded bitstream to reduce the number of frames needed to be transmitted during a FFS or FBS period.

In this article, we address the problem of coding a video sequence by using a binary tree GOP structure so that the resulting video stream can support fast scanning operations very efficiently. We also propose a proximity-based approximation scheme and a modified depth first search (MDFS) algorithm that can further reduce the transmission overhead.

2. Hierarchical Encoding Scheme

2.1. Coding with Ordinary GOP Structure

In a video streaming system, when a client requests for performing fast scan operations, due to the coding dependencies between video frames, the server needs to transmit extra frames to guarantee that all requested frames could be decoded successfully. An example for illustrating the transmission overhead during fast scan operation is shown in Fig. 1. In this example, assume the client requests for performing fast forward scan operation with speed-up factor 4, and the starting frame is an I-Frame. Although there are only 5 frames will be actually displayed, there are total 13 frames needed to be sent and causes a 2.6 times transmission overhead.

![Figure 1](image-url)

Fig. 1. The requested frames (dark grayed frames) and transmitted frames (dark and light grayed frames) for performing a 4x FFS operation.

The random access operation, on the other hand, is on the basis of performing any kind of fast scan operation, when a client requests for random accessing...
a frame, the server needs to transmit a frame with an arbitrary distance from the current frame. For an ordinarily encoded bitstream (see Fig. 2(a)), we can view the coding structure as a completely skewed tree. Suppose all the GOPs in the bitstream have the same size \( n \), the average number of frames to be transmitted for randomly accessing an arbitrary frame for the ordinary encoding scheme is given by

\[
\frac{1}{2} \sum_{i=0}^{n} i = \frac{n+1}{2}
\]

### 2.2. Coding with Binary Tree GOP Structure

Consider Fig. 2(b), where the coding structure for the bitstream is reshaped as a binary tree. The constructing algorithm is illustrated as follows:

**Input:** \( n \) video frames

**Output:** Binary tree GOP structure

1. Select the \( \lceil n/2 \rceil \)-th frame as the I-Frame. Set this frame belong to level \( L = 0 \) and all other frames belong to level \( L + 1 \).

2. Each frame in level \( L \) has both forward and backward predictions striding across half of the remaining successive frames in \( L + 1 \). Set the frames predicted in this step belong to level \( L + 1 \) and the remainder belong to \( L + 2 \).

3. If the number of frames in \( L + 2 \) equals 0, stop the algorithm. Otherwise, let \( L = L + 1 \) and go to Step 2.

For convenience, the size of GOP \( n \) is set to \( 2^k - 1 \), where \( k \) is a certain nature number. Thus the average number of frames to be transmitted in this hierarchical coding scheme is

\[
\sum_{i=0}^{\log_2 n - 1} 2^i \cdot (i + 1) = \sum_{i=0}^{\log_2 n - 1} 2^i \cdot i + n
\]

\[
= \frac{2 - n \log_2 n + 2n \log_2 n - 2n}{n} + 1
\]

\[
= \log_2 n \left( n + \frac{2}{n} \right) - \log_2 n - 1
\]

It is obvious that the hierarchical coding structure is better than the ordinary (linear) coding structure for supporting both of random access and fast scan operations.

The proposed hierarchical encoding scheme can support forward and backward playback for any speed-up factor. To play a bitstream in normal speed, the server follows the depth first search (DFS) order to transmit frames one by one. For example, the transmission sequence for playing a GOP shown in Fig. 2(b) will be: 7, 3, 1, 0, 2, 5, 4, 6, 11, 9, 8, 10, 13, 12, and 14. This transmission will introduce some initial playback delay when the first GOP is requested, and the client needs to buffer some frames for successfully decoding. Fortunately, both of the initial delay time and the client side buffer size will not exceed the half GOP size (e.g. 1/4 seconds in this example, if one second takes 30 frames), and it is considerably to be acceptable for most clients.

![Fig. 2](image_url)

**Fig. 2.** (a) The coding structure of an ordinarily encoded bitstream. (b) The coding structure of a hierarchically encoded bitstream looks like a balanced binary tree.

### 2.3. Proximity Approximation

The proposed binary tree GOP structure can support fast scan operations with high speed-up factors (e.g. large than the GOP size) very efficiently. For the case where speed-up factor is small than the GOP size, we here offer an improvement mechanism that can further reduce the overhead for transmitting the number frames which will not be actually played.

Our approach is based upon the observation: The human visual system is normally not very sensitive to minor change of played scenes; especially when those scenes are changed rapidly. For example, if the desired scene sequence is: 1, 3, 5, 7, and 9. Most of the people cannot tell the difference from this sequence with the sequence 2, 4, 6, 8, and 10, or even the sequence 1, 2, 4, 6, and 9. We therefore apply this observation to the hierarchical encoding scheme and name this scheme as proximity approximation.
2.4. MDFS Algorithm

Now, we propose a modified depth first search (MDFS) algorithm to accomplish the proximity approximation model. Assume the requested frame sequence is

\[ f_1, f_2, f_3, f_4, f_5, \ldots \]

where \( f_i \) represents the frame number of the \( i \)th requested frame, because the same model can equivalently applied on both of forward and backward scan operations, we only discuss the forward scan case here, e.g. \( f_i < f_j \) if \( i < j \).

When frame \( f_i \) is requested for transmission without the proximity approximation, the original depth first search algorithm is performed on the coding tree to find a path from the root (I-Frame) to the node labeled as \( f_i \). On the other hand, if the proximity approximation model is used, we modify the search termination condition as following: when the original DFS algorithm reaches a node \( f_j \) that satisfy the condition

\[ f_i \leq f_j < f_{i+1}, \]

the search procedure is terminated, and returns \( f_j \) as the searching result.

2.5. An Illustrated Example

In this section, we give an example to demonstrate the coding efficiency of the proposed schemes. Consider that the GOP size is 15 and the fast forward scan speed-up factor is 5. The requested frame sequence will be: 0, 5, and 10. If the ordinary encoding structure is employed, we have to transmit total 11 frames (e.g. from 0 to 10). If the proposed structure is applied, the number of transmitted frames will be reduced to 8 (e.g. 7, 3, 1, 0, 5, 11, 9, and 10). Besides, if we apply the binary tree GOP along with our approximation model, the number of transmitted frames can be further reduced to 3 (e.g. 7, 3, and 11).

3. Simulation Results and Discussions

In this section, we give the simulation results of the proposed approaches. Four test sequences (coastguard, Foreman, Hall and Mobile) in CIF format are used for simulation. We fix the GOP size to 15 and encode the test sequences in MPEG-1 format with linear and binary tree GOP structures. The resulted bitstreams are used to simulate the average number of transmitted frames for decoding one request frame and the average PSNR values for various speed-up factors. There are five transmission cases are considered in our experiments.

Case A: Ordinary GOP structure, start playing at the first frame (e.g. an I-Frame)
Case B: Binary tree GOP structure, start playing at the first frame (e.g. a P-Frame)
Case C: Binary tree GOP structure, start playing at the 7-th frame (e.g. an I-Frame)
Case D: Binary tree GOP structure, proximity approximation enabled, start playing at the first frame (e.g. a P-Frame)
Case E: Binary tree GOP structure, proximity approximation enabled, start playing at the 7-th frame (e.g. an I-Frame)

The average number of transmitted frames for decoding a requested frame for various speed-up factors is shown in Fig. 3. When the speed-up factor is a multiple of the GOP size, the average number of transmitted frames for decoding a requested frame will be 1 for Case A, Case C, and Case E, this is due to the fact that all requested frames are I-Frame. However, for Case B and Case D, the average number of transmitted frames for decoding a requested frame will achieve the peak values because all requested frames (P-Frame) have the longest distance to their corresponding I-Frame. The longest distance will be the height of the tree, which is also the maximum number of extra frames to be sent when performing fast scan operations.

![Fig. 3. Estimated number of transmitted frames for decoding a requested frame in various speed-up factors.](image-url)
Table 1. Average number of transmitted frames for various speed-up factors and GOP types.

<table>
<thead>
<tr>
<th></th>
<th>1x</th>
<th>2x</th>
<th>3x</th>
<th>4x</th>
<th>8x</th>
<th>16x</th>
<th>24x</th>
<th>32x</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.00</td>
<td>1.93</td>
<td>2.60</td>
<td>3.00</td>
<td>5.95</td>
<td>6.84</td>
<td>6.54</td>
<td>7.00</td>
</tr>
<tr>
<td>B</td>
<td>1.00</td>
<td>1.47</td>
<td>2.00</td>
<td>2.13</td>
<td>2.79</td>
<td>3.26</td>
<td>3.31</td>
<td>3.70</td>
</tr>
<tr>
<td>C</td>
<td>1.00</td>
<td>1.48</td>
<td>1.80</td>
<td>2.15</td>
<td>2.81</td>
<td>3.21</td>
<td>2.92</td>
<td>3.30</td>
</tr>
<tr>
<td>D</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.20</td>
<td>1.74</td>
<td>2.26</td>
<td>2.31</td>
<td>2.10</td>
</tr>
<tr>
<td>E</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.20</td>
<td>1.73</td>
<td>2.11</td>
<td>2.15</td>
</tr>
</tbody>
</table>

On the other hand, when performing fast scan operations, our PSNR results are also similar to the ordinary GOP (see Fig. 4). The slight degradation or enhancement of quality in Case B and Case C is caused by the reconstruction of the GOP structure. If we apply the approximation scheme, the quality level is degraded for about 0.5 dB, but the number of transmitted frames can be significantly reduced.

Fig. 4. Estimated PSNR for various speed-up factors.

It can be easily seen that both the estimated number of frames to be sent and bitrate are much fewer than that of the ordinary GOP. Comparing with the dual bitstream scheme [5], where the estimated number of transmitted frames is approximately 2.71. Our results are also close to this value, even better than it if the approximation approach is employed. Moreover, we do not have to encode two bitstream in the encoder side. Certainly, the average bit-rate while performing VCR operations are significantly decreased as well.

The most important question we have to answer here is that the longer predicted distance may cause quality degradation as well as increase the coding bitrate. We conducted several experiments on this issue and the results are shown in Table 2, which shows the comparison in average PSNR and bitrate result between the ordinary GOP and the binary tree GOP (Case D and Case E are ignored here because the proximity approximation model is useless here).

Table 2. PSNR and bitrate comparisons between the ordinary GOP and the binary tree GOP (Case D and Case E are ignored here because the proximity approximation model is useless here).

<table>
<thead>
<tr>
<th>GOP type (k=4)</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg. PSNR (dB)</td>
<td>Avg. bitrate (MB/s)</td>
<td>Avg. PSNR (dB)</td>
</tr>
<tr>
<td>Coastguard</td>
<td>30.69</td>
<td>30.41</td>
<td>30.41</td>
</tr>
<tr>
<td>Foreman</td>
<td>32.84</td>
<td>32.63</td>
<td>32.6</td>
</tr>
<tr>
<td>Hall</td>
<td>34.17</td>
<td>33.96</td>
<td>33.95</td>
</tr>
<tr>
<td>Mobile</td>
<td>26.67</td>
<td>26.38</td>
<td>26.37</td>
</tr>
</tbody>
</table>

4. Conclusions

In this article, we proposed a novel approach to reorganize the ordinary GOP structure to support fast scan operations for video streaming applications. We employ the binary tree structure along with a proximity approximation model. Our approach can be applied both for FFS and FBS with any speed-up factors even for the normal playback. Our future work includes enhance the frame selection scheme and explore other hierarchical encoding structure (e.g. m-ray tree) to find other possibilities to reduce the transmission and storage overhead to support VCR functionalities economically.

References