Improved Two-Layer Coding Schemes for Motion Picture Sequences

Shang-Pin Chang, Tsorng-Yang Mei and Hsueh-Ming Hang
Dept. of Electronics Eng.,
National Chiao-Tung University
Hsinchu 300, TAIWAN, ROC
886-35-712121 ext. 3298
hmhang@cc.nctu.edu.tw

ABSTRACT

This paper proposes several improved schemes on the two-layer codec. In a two-layer coder, the packets produced by the base layer are set to the high priority and those produced by the second layer are low. Three improved schemes on the base layer are proposed. The base layer bits and/or the total bits have been significant reduced by these improved algorithms. In addition, the base images have been also remarkably enhanced.

INTRODUCTION

The flexible bit rate capability of ATM (Asynchronous Transfer Mode) networks matches well the variable bit rate nature of video compressors. However, ATM networks also introduce new problems that are not encountered in the constant bit rate channels, such as packet loss. Multi-layer variable bit rate algorithms are proposed to meet the ATM network requirements. They lead to a rather natural priority assignment. When network is in congestion, the low priority cells are discarded first so that the high priority cells that contain more important information are retained. In addition, a two-layer coding structure with spatial partition can provide separate components for multi-grade services [1]. The most essential component is assigned the *base* layer, and the rest, *enhancement layer*. The data cells from the base layer are naturally given the high-priority.

We will describe a conventional two-layer coding system which serves as the fundamental structure and then three improved base layer coding schemes for enhancing the picture quality. The layer signal separation is performed in the spatial domain.

CONVENTIONAL DECIMATION

The block diagram of the first scheme, two-layer coding with conventional decimation, is showed in Figure 1. The original images are in CIF format (288 lines by 352 pels). After they are lowpassed and decimated, their sizes are scaled down from CIF to QCIF (144 lines by 176 pels). These QCIF images are the inputs to the base layer — an interframe RM8 video coder — and the output codes of this base layer are assigned higher priority. Decoding these codes, the QCIF images can be reconstructed and then upsampled to CIF size. Since these CIF pictures are bilinear interpolated from QCIF compressed images, they are called the *interpolated* pictures. At enhancement layer, a modified JPEG image coder is proposed to encode the residual images generated by subtracting the interpolated pictures from originals. The output codes are the enhancement layer data.

The reason of adopting this two-layer architecture is that it separates the information into different levels of importance. The images transmitted at base layer are decimated by a factor of 2 in both horizontal and vertical directions. The base layer images contain the most essential information of the original images and thus retain most of the characteristics of the originals. Therefore, motion compensation of RM8 coding algorithm performs well and the interpolated pictures are similar to the originals. However, because of the interframe motion compensation technique used at the base layer, lost packets would dagrade the decoded image quality severely. Hence, the base layer information should have very high priority; that is, very low loss rate.

The residual images transmitted at the enhancement layer are only to be added to the interpolated pictures to generate *high-resolution (HR)* pictures. Partial loss of this add-on information often only degrades slightly the reconstructed images quality. This higher tolerance of data loss allows their information be transmitted with lower priority.

Scheme 1 can be improved by applying a low-pass filter to the original picture before decimation. This is because in the decimation process the high-frequency components of the original image are aliased into the low-frequency band. The interpolated pictures are hence degraded. Therefore, pre-filtering using a half-band low-pass filter

improves the quality of the interpolated pictures.

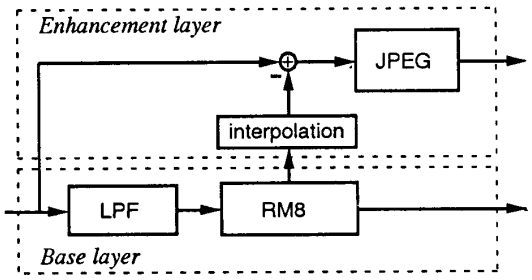

Figure 1: Block diagram of a conventional two-layer system (Scheme 1)

PROGRESSIVE RESAMPLING

Stationary-background video

Although the smallest step size of RM8, 4, is used at the base layer, the reconstructed interpolated pictures cannot achieve the desired quality. This quality loss comes mostly from the fact that 3/4 pixel values of interpolated pictures are obtained by interpolation and thus generate large amount of errors. If we can improve the quality of interpolated pictures without increasing their bits, we could reduce significantly the bit number of the enhancement layer. Therefore, we propose a modified decimation technique that, when time progresses, it accumulates more than 1/4 of the compressed pels of the original pictures. Thus, the interpolated picture quality is significantly improved, and therefore, the interpolated pel differences are much smaller. Consequently, fewer bits are necessary to be transmitted at the enhancement layer. This scheme will be called Scheme 2.

In order to preserve more original pixel values, the conventional decimation is replaced by periodic resampling. The stationary parts of images, such as background, are subsampled (decimated) according to a periodic pattern up to four consecutive decimated images. In this case, after four frame transmission, every pel in the stationary regions is coded and reconstructed to form an original resolution image (rather than an interpolated image). The details of this scheme are described below.

If a 16 by 16 image block stays stationary for four frame-time, during the first frame-time one out of every four (2 by 2) pels is grouped together to constitute an 8 by 8 block. Then, DCT and quantization are applied to this 8 by 8 block. During the next frame-time, the every second pel in this four pel pattern is grouped together and transform-coded. Hence, after four frame time, every single pel of this 16 by 16 image block is coded with an accuracy controlled by the quantizer step size. In general, this quantization error is much smaller than the interpolation error. In fact, this reconstructed block is often indistinguishable from the original one when the quantization step size is small enough. Therefore, encoding the residual image of this block generates very few bits. On the other hand, the increase of base layer bits is insignificant since the interframe differences that are coded and transmitted are fairly small.

This scheme employs periodic resampling and stationary blocks detection. The detection process retains one CIF size frame and three QCIF frames in memory. The CIF frame contains the latest original picture for comparing to the current picture to detect stationary blocks. A *stationary duration table* records the stationary frame-times of each blocks up to the current frame. The last three decimated (coded) images are used to reconstruct the current interpolated picture.

Non-interlaced panning video

In Scheme 2, the stationary property of image background is utilized to enhance the quality of interpolated pictures. After receiving several consecutive images, the receiver can reconstruct the stationary regions by placing the coded original resolution pels in their locations rather than by conventional bilinear interpolation. The drawback of Scheme 2 is the very strict condition for the regions to be considered stationary. If in two consecutive images the same object are doing translational movement, it will not be classified as a stationary region. However, this occurs frequently, for example, in an image sequence with camera panning. Such being the case, Scheme 2 fails to improve the quality of interpolated pictures.

A modified version of Scheme 2, Scheme 3, is proposed to compensate translational movement for non-interlaced videos. It finds the best matched blocks of present blocks in the previous frame. If the motion compensated difference blocks also satisfy the stationary criterion defined in Scheme 2, they are classified as translational block. The translational blocks are grouped into several panning groups, each contains blocks of the same motion vector. These motion vectors and the group indexes of panning blocks are coded; then, the decoders are able to reconstruct better interpolated pictures using the same procedure in Scheme 2 but with an additional trace-back step that finds the aggregate motion vectors of the original resolution pels in the preceding frames.

Unlike Scheme 2, Scheme 3 contains irregular steps in generating the downscaled images from the original images and in generating the upscaled images from the received base-layer frames. Decimated images are generated by picking up a pel from every four (2 by 2) pels not in a regular periodic order but in the order decided by motion. This is partially due to the following pan-

ning compensation procedure that may disturb the resampling orders described in Scheme 2. Thus a pixel source table in CIF size is needed to remember the original frame index of each pixel up to a 4-frame time duration. It should be noted that this table can be reconstructed by panning vectors at the receiver, therefore; no bits are needed to be transmitted.

Each number in the pixel source table indicates an index of a frame from which the corresponding pel can be copied with the help of an aggregate motion vector. The possible values in this table are '1','2','3','4' and '0'. The first four numbers represent the latest four frame indexes of a sliding window and the last one, '0', denotes none of the four reference frames can produce the current pel. After four frames being transmitted, a nonzero number in the pixel source table indicates that it is copied from a more recent frame (especially, '0' represents its original pel does not exist in the reference frames). Thus the table can tell us where are those pels coping from if they are moved from the previous frames. For those pels that are indicated by '0' on this table, they have the highest probability of being transmitted (that is, being the elements of the QCIF size frame). In other words, each pel of a QCIF size frame is chosen from its corresponding 2 by 2 pels of CIF size frame in accordance with the priorities of the four pels. The '0' pels have the highest priority. They are followed by the '1' pels and so on. If two or more pels have the highest priority, one of them is picked according to the raster scan order. After a pel of every 2×2 block in a CIF size frame is selected, its location in the pixel source table is recorded with the current sliding frame index.

When generating the to-be-transmitted decimated pictures, the pixel source table provides the source frame indexes of the transmitted original resolution pels which can be found in the most recent four frames. Therefore, by looking up the group indexes table of panning blocks and the panning group motion vectors table of these four frames, the aggregate motion vectors of these transmitted pels can be calculated by concatenating motion vectors between every two consecutive frames. Then the transmitted-pel picture of the current frame is constructed by coping each transmitted pel from the previous transmitted decimated frames according to the associated concatenated motion vector.

The steps for tracing back an available pel are illustrated by Figure 2. The original resolution pel of the left-top most pixel of frame 4 can be copied from the reference QCIF frame 1 (that is, in fact, 3 frames away from the current QCIF frame) as denoted by the pixel source table. We first follow the motion vectors between frames 3 and 4 and trace back to the pel location on frame 3 of this current pel. Then, we follow the motion vectors between frames 2 and 3 and find the current pel location on frame 2. Finally, we locates the current pel location in frame '1' from the motion vectors between frames 1 and 2. In this example, we need to trace back 3-frame time to obtain this pixel. Thus an aggregate motion vector through 3-frame time duration is obtained by concatenating the panning vectors along the trace of the current pel in the past four frames. Finally, the location of the current pel, (X, Y), in frame 1 can be calculated by adding up this aggregate motion vector and the coordinate of the current pixel. And the current pel can be retrieved from the reference QCIF frame 1 at $(X/2, Y/2)$, because the coordinates of a QCIF frame equal to half of their counterpart in the CIF frame. The not-transmitted pels are indicated by 'X' in Figure 2. They are produced by bilinear interpolating the transmitted pels.

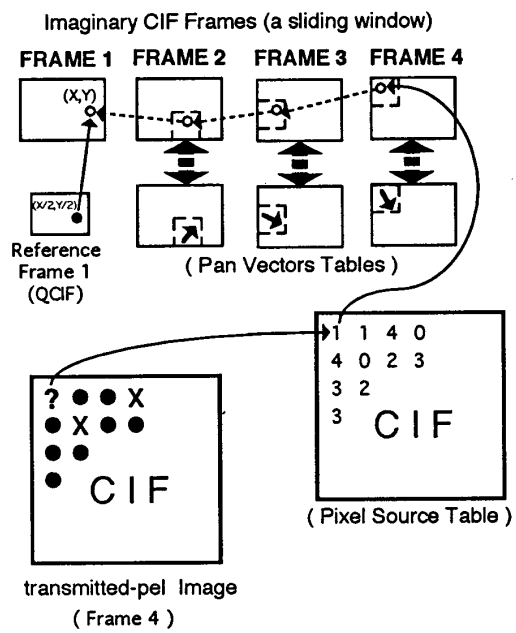

Figure 2: An illustration showing the transmitted-pel image reconstruction in Scheme 3

Interlaced panning video

Because the adjacent fields of an interlaced frame are related both temporally and spatially, consecutive frames may be different for objects with non-uniform motion. Therefore, it may not be effective to simply use the frame-based motion compensation for an interlaced video. We modified the motion estimation and the spatial interpolation algorithms of Scheme 3 to be field-based ones. This new version, Scheme 4, is thus particularly suitable for encoding interlaced videos.

SIMULATION RESULTS

Figure 3 shows the statistics of bits-per-frame and PSNR of three image sequences coded by all the above schemes. It is clear that the quality of interpolated pictures has been improved remarkably by using the progressive resampling algorithm (Scheme 2, 3, and 4) for stationary-background video ("Swing" sequence), and hence the total bit number is significantly reduced. And note that Scheme 3 and Scheme 4 have about the same performances for the "Swing" sequence but they are much superior than Scheme 2 for encoding images with a large amount of non-interlaced panning ("Chairlady" sequence). Moreover, Scheme 4 performs very well even on the interlaced panning images ("Harbour" sequence) while Scheme 3 fails.

CONCLUSIONS

In this paper, three modified 2-layerschemes generally can improve the performance of the conventional scheme quite significantly. They either use fewer bits to transmit the same or more information, or achieve much higher picture quality with a little more additional bits. The most noticeable improvement is offered by the progressive resampling approach. The original pixels are transmitted progressively in the base layer to increase the interpolated picture quality. Both non-interlaced and interlaced video sources have been considered and the specific schemes for handling them are designed.

ACKNOWLEDGEMENT

This research has been supported by a grant from Telecommunication Laboratories, Ministry of Transportation and Communications, Taiwan, ROC.

REFERENCES

[1] M. Ghanbari, "Two-Layer Coding of Video Signals for VBR Networks", IEEE Journal on Selected Areas in communications, Vol.7, No.5, pp.771-781, June 1989.

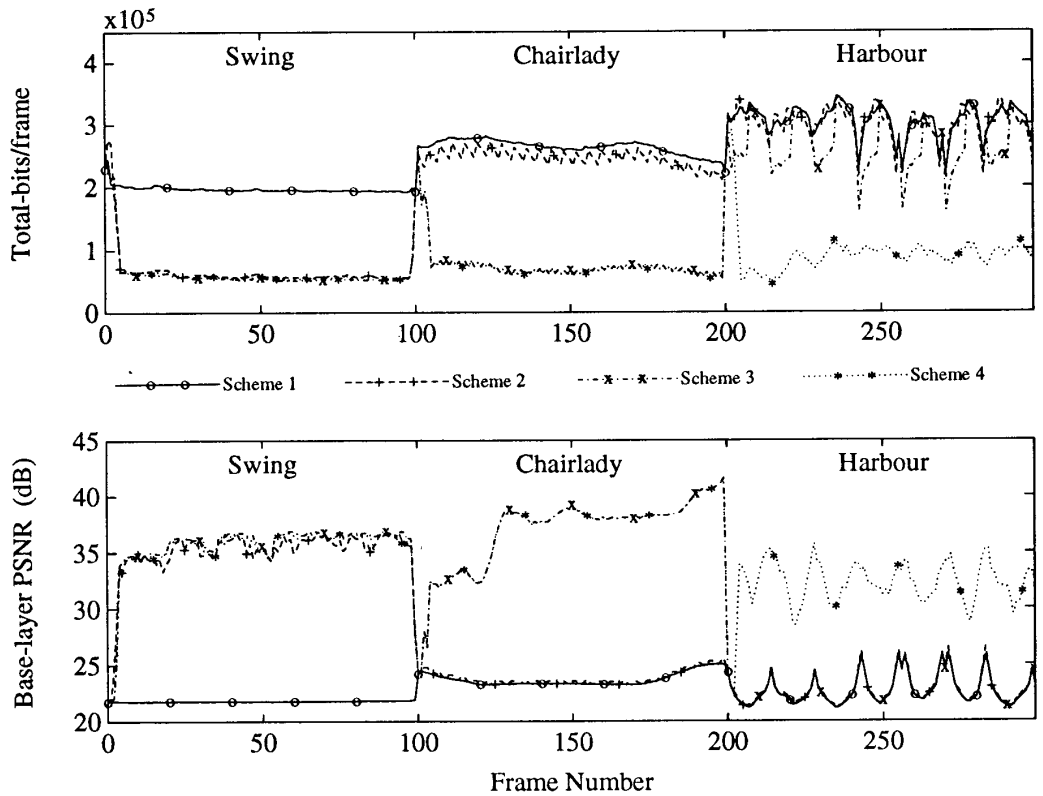

Figure 3: Statistics of bits and PSNR of three specified sorts of videos coded by the four proposed schemes