PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Error resilient video coding using virtual reference picture

Zhang, Guanjun, Stevenson, Robert

Guanjun Zhang, Robert L. Stevenson, "Error resilient video coding using virtual reference picture," Proc. SPIE 5685, Image and Video Communications and Processing 2005, (14 March 2005); doi: 10.1117/12.587306

Error Resilient Video Coding Using Virtual Reference Picture

Guanjun Zhang and Robert L. Stevenson
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556, USA

ABSTRACT

Due to widely used motion-compensated prediction coding, errors propagate along decoded video sequence and may result in severe quality degradation. Various methods have been reported to address this problem based on the common idea of diversifying prediction references. In this paper, we present an alternative way of concealing the references pictures errors. A generated virtual picture is used as a reference instead of an actual sequence picture in the temporal prediction. The virtual reference picture is generated in a way to filter damaged parts of previously decoded pictures so that the decoder can still get a clean reference picture in case of errors. Coding efficiency is effected due to the fact that the virtual reference is less correlated to the currently encoded picture. The simulations on H.264 codec have shown quality improvement of the proposed method over intra-coded macroblock refreshment. It can be used on any motion-compensated video codec to combat channel errors.

1. INTRODUCTION

Error resilience become an important feature of video transmission over error-prone networks. Generating error resilient video bit streams has become an intensive research area and many solutions has been reported to address this problem.

In conventional motion-compensated prediction video coding, errors in a single picture may cause mismatch between encoder and decoder so that the effect of errors may propagate along decoded video sequence and thus result in severe quality degradation. A common way is using independently decodable video coding structures in different levels from macroblock to sequence. The picture level refreshment is commonly found as I frame in all coding standards. In sequence level, there is multiple description coding, such as Video Redundancy Coding(VRC) in H.263+. [1] The refreshment using intra coded macroblocks has been proven to be an effective way to combat errors due to packet loss or bit errors. The side effect of using intra coded macroblocks is loss in coding efficiency.

A another basic idea of concealing error effect in inter coded video stream is to use robust prediction reference so that the visual quality variation is alleviated. Girod et. al. diversify the references of motion-compensated prediction to improve the coding efficiency[2],[3] while Wang et. al. found out it can also function against errors,[4] as long as a balance between efficiency and robustness is achieved. Multiple references of motion-compensated prediction has been adopted into H.263 and H.26L as an important tool for robustness as well as higher coding efficiency [5].[6] Reference of current pictures are distributed in several previously coded pictures. If each currently coded block use one reference block, the situation is not different from conventional prediction since a block is damaged when its reference has errors. If more than one reference blocks are searched, it may cost a large amount of additional time and power. In [7] the authors find out that, although multiple reference are used, the actual references are unevenly located in the temporally previous picture and a fast motion estimation algorithm is then developed.

In error concealment capable video codecs, the decoders always first try to conceal errors in current decoded pictures before using them as references. If back channels are not available or timely inapplicable, the decoders usually directly copy the same area from another picture. If more time and complexity is affordable, a space interpolation or motion-compensated interpolation can be applied to yield more accurate reconstructions.

To further improve error concealment, methods of flexible slice structures are standardized in MPEG-4 AVC/H.264. It basically distributes macroblocks from different areas of the pictures into a slice structure. When a slice is crushed, the benefits are two-fold: visual quality is less effected since the macroblocks are scattered in the picture; and space

Tel: (574)631-8308 Fax: (574)631-4393 Email: {gzhang, rls}@nd.edu

interpolation is more efficient because damaged macroblocks have more opportunities to be surrounded by error-free macroblocks from other slice. Although these concealment steps utilize clean content from earlier decoded pictures, the damaged areas can only be partially recovered in most cases. The mismatch problem still exists, if the concealed areas do not happen to completely match the original ones.

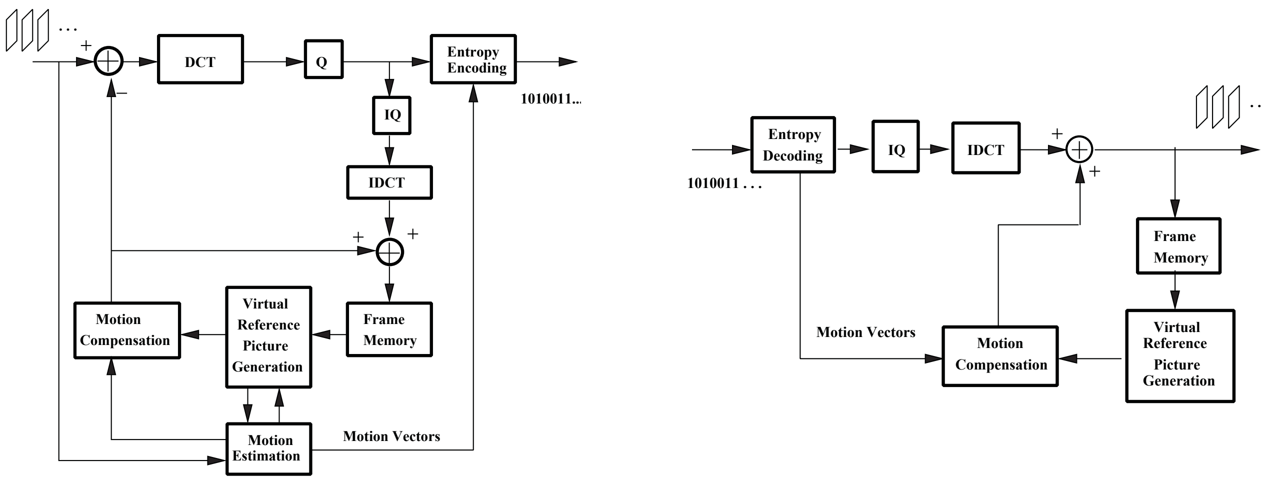
In this paper, we present a video coding system which has the capability to combat the error propagation in a damaged video sequence. A virtual reference picture(VRP) is generated by using a non-linear filter on several real pictures in both encoder and decoder, and then used to predict the incoming picture. It become an novel solution to the mismatch problem in terms of more chances to completely recover the damaged areas than the above systems.

The rest of this paper is organized as follows. In section 2, we will introduce the video coding system using virtual reference pictures and discuss its error resilience capability. In section 3, we will describe the simulation environment. Simulation results will be presented and discussed in section 4, which is followed by the conclusions.

2. VIRTUAL REFERENCE PICTURE

In today's popular video codecs, motion-compensated prediction uses earlier decoded pictures as references, which are actually displayed after decoding. If a back channel or retransmission is not available to the system, a damaged picture becomes a damaged reference at the decoder. We argue that it is not necessary to use a real video frame as a prediction reference. A virtual picture generated from real pictures can also be use as the reference, as long as it has appropriate error resilience properties and acceptable price in the sense of efficiency loss.

We propose a video coding system, as illustrated in figure 1. The encoder generates a virtual picture by applying a nonlinear filter on several previously coded pictures. It then uses the virtual picture to predict the incoming picture. A same filter is used in the decoder to generate matched reference picture as in the encoder. We assume that there is no back channel available.



(a) Encoder (b) Decoder

Figure 1. A video coding system using virtual reference pictures.

In an error free environment, both encoder and decoder use identical virtual reference pictures. In case of errors, the decoder first conceals the errors in a picture by spatial interpolation or simple temporal interpolation, such as a copy from the same areas in a previously decoded picture. Then the virtual reference picture is yielded by using the concealed picture and several other decoded pictures.

If we consider the concealed picture as a copy of the encoded picture which is damaged by additive noise, the filter shall have the capability to conceal the noise from this picture and generate a virtual picture mostly using information from other clean pictures. When this virtual picture is used to predict the next incoming picture, the effect of errors in the next picture is further constrained since its reference has more correct data compared to the damaged picture.

In this work, we assume the difference between the concealed picture and the error-free picture is equivalent to additive white noise for simplicity. We will only use and test the nonlinear median filter to generate the virtual reference pictures due to its capability of conceal white noise. The virtual reference picture is generated from three previously decoded pictures. Every pixel value in the virtual reference picture is obtained by median filtering the three pixels, which are from the same position of the three decoded pictures, respectively. If we only use forward prediction, the dependence of the real sequence pictures and the virtual reference pictures are illustrated in figure 2. the solid arrows represent the prediction directions and the dash arrows represent contributions of real pictures to the virtual reference pictures. Note that the first two pictures of a video sequence still use the previous real picture as their reference, respectively.

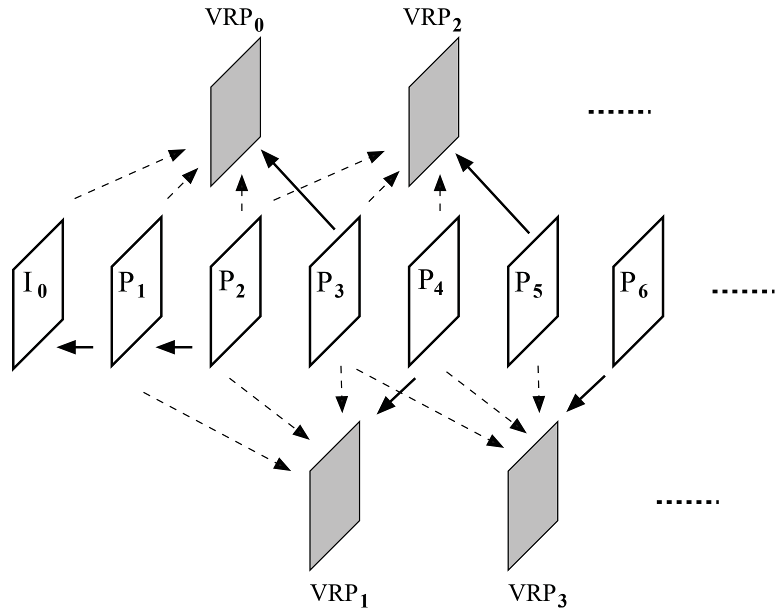


Figure 2. Dependence of pictures in the virtual reference video coding system.

Due to the lower correlation between the virtual reference picture and the incoming picture, coding efficiency in the proposed video coding system will be lower than the conventional coding system. In section 4, we will show that the bit rate increment is acceptable and comparable to that of the macroblock refreshment in most tests.

3. SIMULATION SETUP

The proposed video coding system is built based on MPEG-4 AVC/H.264 reference software JM9.0.[8] Simulations are run on both Virtual Reference Picture(VRP) codec and the original H.264 codec, and their results are compared. The error concealment of a damaged picture involves spatial and temporal interpolations as developed in.[9] The concealed picture along with two more previously decoded pictures are used to generate the virtual reference picture.

All pictures of a test sequence are coded in forward prediction frame mode except the first I picture. All valid types of motion estimations and predictions are allowed as in H.264 baseline. At this stage, only intra macroblock refreshment is used in both codecs as the error control method in order to compare their performance in error-prone environments. Test video sequences are encoded by both VRP and H.264 encoder using the same H.264 baseline parameters except settings for how many intra macroblock per picture. The VRP encoder use the same or half number of intra macroblocks than the H.264 encoder. In H.264, only one reference picture is use.

The noisy channel is assumed to be the binary symmetric channel as in many wireless applications. Random bit errors are added to the coded bit streams to simulate the situations when lower level protection, i.e., channel coding, fails. These errors may damage header or data information and cause termination of decoding a slice. Information in the slice after the errors are considered lost. For simplicity, we set each slice to have the same number of macroblocks in order to synchronize displaying of next slice.

The first test is to demonstrate that the VRP system is able to conceal errors of a decoded picture and constrain the propagation of errors. Uniformly distributed random bit errors are generated and added to both bit streams. PSNR values of each decoded or concealed picture in the sequences are recorded and compared. Rate distortion optimization is set to work so that compared sequences have the same PSNR. Coding efficiency is compared in terms of bit stream length.

The second test is to statistically show that the VRP system has more protection to the video streams than the intra macroblock refreshment. This test collects PSNR values of decoded sequences at different bit error rates(BER) and error patterns. At each interested BER point, the first test is repeated one hundred times under different errors and yield an averaged PSNR value. Then a trend of video quality drop along with increased BER can be drawn. Comparing curves of two system will bring the conclusion of error resilience capability.

4. RESULTS AND DISCUSSION

Error resilient capability of the VRP video coding system is first illustrated. The test sequence *mobile* contains 30 pictures and is in 4:2:0 SIF format(352×240). Quantization parameters are set to 28 for all pictures in both encoders. In both sequences, there are 10 intra macroblocks in each H.264 inter picture and 5 in each VRP inter picture. Figure 3 shows a typical example of the PSNR values of the damaged sequences. Between picture No.7 and No.9, The proposed VRP sequence only experiences a short period of quality drop and recover quickly, while the H.264 sequence needs a very long time to recover from errors in the single picture No.8. The same thing happens to pictures between No.12 and No.16. This results in a higher average PSNR of VRP sequence(27.77dB) than H.264 sequence(25.05dB). Pictures from No.7 to No.9 of VRP sequence are displayed in figure 4, which shows that the part in the middle of the calendar is damaged and concealed. The PSNR drops that happen earlier in both sequences are due to some errors that can not be concealed by the current algorithm implementations in the reference software JM9.0.

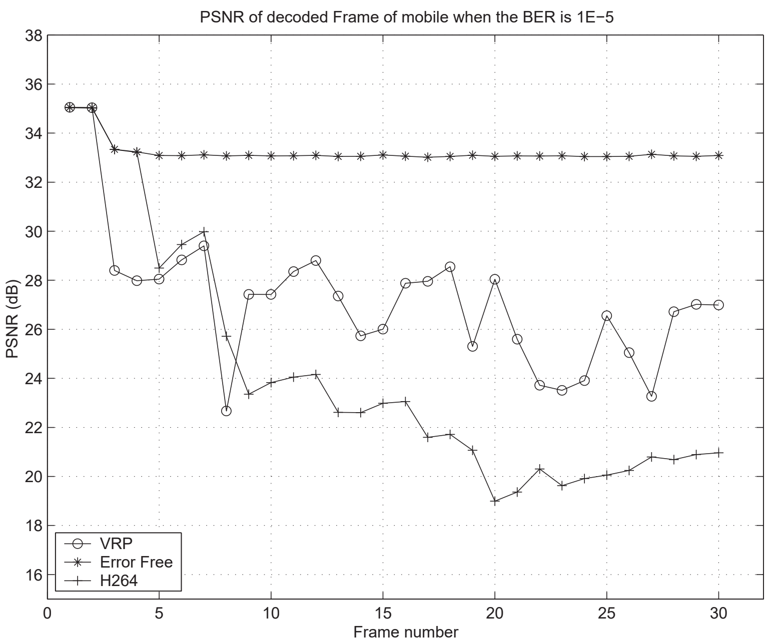


Figure 3. PSNR value comparison of sequence mobile.

At a frame rate of 30 frame per second, the average bit rate of each P frame in *mobile* is 10.9 KByte/s from the H.264 encoder and 16.2 KByte/s from the VRP encoder. The overhead of 48.7% largely dues to the fact that complicated context and details of a real picture is filtered by the median filter. Because the VRP bit stream is longer than the H.264 bit stream, it also encounters more errors. This makes the VRP curve have more pits than the H.264 curve.

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.