RANDOM ACCESS USING ISOLATED REGIONS

Miska M. Hannuksela[1], Ye-Kui Wang[1], and Moncef Gabbouj[2]

[1]Nokia Mobile Software, Tampere, Finland
[2]Tampere University of Technology, Finland

ABSTRACT

Random access is a desirable feature in many video communication systems. Intra pictures have been conventionally used as random access points, but correct picture content can also be recovered gradually within a range of pictures starting from a non-intra random access point. This paper proposes the use of the isolated regions technique for gradual decoder refresh and presents how the proposed method can be used in the upcoming ITU-T Recommendation H.264, also known as MPEG-4 Part 10 or Advanced Video Coding. The presented simulations reveal that the proposed method outperforms intra-picture-based random access points in error-prone network conditions. It is also shown that the proposed method is more flexible and suits packet-based transmission better compared to progressively located intra-coded slices.

1. INTRODUCTION

1.1. JVT Coding Standard

The Joint Video Team (JVT) of ITU-T and ISO/IEC is actively working on a new standard based on an earlier ITU-T standardization project called H.26L. The resulting standard [1] is called ITU-T Recommendation H.264 or ISO/IEC International Standard 14496-10 (MPEG-4 Part 10) and is referred to as the JVT coding standard in this paper.

A coded JVT sequence consists of Network Abstraction Layer (NAL) units. A NAL unit is an atomic element that can be framed for transport and parsed independently. Each NAL unit has a specific type, which can be a coded slice, a sequence parameter set, a picture parameter set, or a supplemental enhancement information (SEI) message, among other things.

Some coding parameters in a NAL unit of one type depend on coding parameters of a NAL unit of another type. In particular, the following dependency hierarchy relates to coded slices: A sequence parameter set can be decoded independently. A picture parameter set contains parameters that remain unchanged within a picture, and it refers to a certain sequence parameter set. A coded slice consists of a slice header and slice data and refers to a certain picture parameter set.

The JVT coding standard enables storage of multiple reference pictures for inter prediction (a.k.a. motion compensation) and selection of the used reference picture on picture segment or macroblock basis. A particular type of an intra picture, called an instantaneous decoder refresh (IDR) picture, is specified in the JVT coding standard. No subsequent picture can refer to pictures that are earlier than the IDR picture in decoding order. A particular NAL unit type is dedicated to IDR pictures in order to enable their detection easily.

Macroblocks can be grouped into more than one slice group. A slice group can contain any macroblock location, including non-adjacent macroblocks. A slice consists of at least one macroblock within a particular slice group in raster scan order. The allocation of macroblock locations to slice groups is specified in the picture parameter set.

Supplemental enhancement information is not necessary to decode sample values correctly. SEI messages may help display the decoded pictures correctly or conceal transmission errors, for example.

1.2. Random Access

Random access refers to the ability of the decoder to start decoding a stream at a point other than the beginning of the stream and recover an exact or approximate representation of the decoded pictures. A random access point and a recovery point characterize a random access operation. The random access point is any coded picture where decoding can be initiated. All decoded pictures at or subsequent to a recovery point in output order are correct or approximately correct in content. If the random access point is the same as the recovery point, the random access operation is instantaneous; otherwise, it is gradual.

Random access points enable seek, fast forward, and fast backward operations in locally stored video streams. In video on-demand streaming, servers can respond to seek requests by transmitting data starting from the random access point that is closest to the requested destination of the seek operation. Switching between coded streams of different bit-rates is a method that is

0-7803-7750-8/03/$17.00 ©2003 IEEE

used commonly in unicast streaming for the Internet to match the transmitted bitrate to the expected network throughput and to avoid congestion in the network. Switching to another stream is possible at a random access point. Furthermore, random access points enable tuning in to a broadcast or multicast. In addition, a random access point can be coded as a response to a scene cut in the source sequence or as a response to an intra picture update request.

Conventionally each intra picture has been a random access point in a coded sequence. The MPEG-2 video coding standard allows the use of group of pictures (GOP) headers in front of intra pictures in decoding order. The GOP header includes a flag to signal if the GOP is closed or open. In a closed GOP, no pictures following the first intra picture of the GOP are predicted from earlier pictures in decoding order. In an open GOP, the GOP may include B pictures that are predicted from the first intra picture of the GOP and from the latest reference picture preceding the first intra picture in decoding order.

Due to the introduction of multiple reference pictures for inter prediction, an intra picture may not be sufficient for random access in the JVT coding standard. For example, a decoded picture before an intra picture may be used as a reference picture for inter prediction after the intra picture. Therefore, IDR pictures are used as random access points instead of ordinary intra pictures.

Gradual decoder refresh (GDR) refers to the ability to start the decoding at a non-intra picture and recover decoded pictures that are correct in content after decoding a certain amount of pictures. GDR can be implemented by coding intra slices "progressively" so that their location changes from one inter picture to another as mentioned in [2]. For example, one macroblock row per picture from top to bottom may be coded as an intra slice.

Some reference pictures for inter prediction may not be available between the random access point and the recovery point, and therefore some parts of decoded pictures in the gradual decoder refresh period cannot be reconstructed correctly. However, these parts are not used for prediction at or after the recovery point, which results in error-free decoded pictures starting from the recovery point. Therefore, in addition to coding every macroblock progressively in intra mode, inter prediction of macroblocks that were previously intra-coded must not refer to any sample position that has not been intra-coded in the same refresh period. Otherwise, the picture is probably not completely correct in content, when all macroblock locations have been intra-coded as demonstrated in [3]. In order to help decoders conclude when a completely correct picture is achieved in GDR, the JVT coding standard includes the random access point SEI message that signals the random access point and the recovery point explicitly. Consequently, the random access point SEI message can also be used to implement functionality similar to the open GOP feature of MPEG-2.

It is obvious that the implementation and use of GDR is more cumbersome compared to instantaneous decoder refresh. However, GDR provides a couple of benefits, including smooth bitrate and error resilience as explained below. Certain applications, such as broadcast video transmission, require frequent random access points. If the desired random access point frequency is greater than the scene change frequency of a sequence, the use of intra pictures as random access points is likely to increase fluctuations in the instantaneous bitrate and in the coded data buffer occupancy level of the codec. Fluctuations in the instantaneous video bitrate can be handled by allowing bursty transmission bitrate or by enlarging the required initial buffering delay in the receiver. These complications can be avoided with GDR. Moreover, if the applied communication network is error-prone, forward error control methods [4] are commonly used to enable the receiver to recover or conceal corrupted or lost data. One of the most common forward error control methods is the family of loss-aware macroblock mode selection algorithms [5]. Encoders use these algorithms to stop temporal error propagation according to expected or known network conditions. As the use of both the loss-aware macroblock mode selection algorithm and the GDR algorithm result into intra-coded macroblocks, it has been shown that the algorithms can be combined efficiently [6].

In many packet-oriented transport systems, the optimal size of a packet and a slice in bytes is a function of expected packet loss rate [7]. It is beneficial to encapsulate one slice into one packet, because if a slice is split to multiple packets, a loss of one of these packets prevents decoding of the whole slice. If GDR is implemented with intra-coded slices, the slice size is fixed in number of macroblocks but varying in number of bytes. Moreover, the position and size of the slices adjacent to the intra-coded slices depend on the position and size of the intra-coded slices. Consequently, the sizes of the slices are likely to be non-optimal and the slice loss rate may grow compared to optimal size selection of slices.

This paper proposes the use of isolated regions in GDR. Isolated regions allow the use of optimal slice sizes in bytes, enable bitrate and computational scalability in the random access operation, and provide a variety of refresh patterns. Section 2 of the paper introduces the isolated regions coding technique. Section 3 summarizes how the isolated regions technique is implemented in the JVT coding standard. The proposed method for GDR is presented in section 4 in the context of the JVT coding standard. Section 5 presents the performed simulations, and section 6 concludes the paper.

2. FUNDAMENTALS OF ISOLATED REGIONS

An isolated region in a picture can contain any macroblock and a picture can contain zero or more isolated regions that do not overlap. A leftover region is the area of the picture that is not covered by any isolated region of a picture. When coding an isolated region, all predictive coding within the same coded or decoded picture, herein referred to as in-picture prediction, is disabled across its boundaries. A leftover region may be predicted from isolated regions of the same picture.

A coded isolated region can be decoded without the presence of any other isolated or leftover region of the same coded picture. It may be necessary to decode all isolated regions of a picture before the leftover region. An isolated region or a leftover region contains at least one slice.

Pictures, whose isolated regions are predicted from each other, are grouped into an isolated-region picture group. An isolated region can be coupled with a corresponding isolated region in each earlier picture within the same isolated-region picture group. An isolated region can be inter-predicted from the corresponding isolated region within the same isolated-region picture group; whereas, inter prediction from other isolated regions is disallowed. A leftover region may be inter-predicted from any isolated region. The shape, location, and size of coupled isolated regions may evolve from picture to picture in an isolated-region picture group.

3. ISOLATED REGIONS IN THE JVT CODEC

Coding of isolated regions in the JVT codec is based on slice groups. The mapping of macroblock locations to slice groups is specified in the picture parameter set. The JVT coding syntax includes efficient methods to code certain slice group patterns, which can be categorized into two types, static and evolving. Static slice groups stay unchanged as long as the picture parameter set is valid; whereas, evolving slice groups can change picture by picture according to the corresponding parameters in the picture parameter set and a slice group change cycle parameter in the slice header. The evolving slice group patterns include horizontal wipe, vertical wipe, box-in, and box-out. The evolving patterns are especially suited for coding of isolated regions and GDR and are described in more details in the sequel.

An evolving slice group is specified by indicating the scan order of macroblock locations and the change rate of the size of the slice group in number of macroblocks per picture. Each coded picture is associated with a slice group change cycle parameter (conveyed in the slice header). The change cycle multiplied by the change rate indicates the number of macroblocks in the first slice group. The second slice group contains the rest of the macroblock locations. Fig. 1 shows an example of the first five change cycles of the first slice group of the box-out type with a change rate of 12 macroblocks.

Fig. 1. Example of an evolving isolated region.

In-picture prediction is always disabled across slice group boundaries because slice group boundaries lie in slice boundaries. Therefore, each slice group is an isolated region or a leftover region.

Each slice group has a unique identification number within a picture. Encoders can restrict the motion vectors so that they only refer to the decoded macroblocks belonging to slice groups having the same identification number as the slice group to be encoded. Encoders should take into account the fact that a range of source samples is needed in fractional pixel interpolation and all source samples should be within a particular slice group.

4. GDR USING ISOLATED REGIONS

Evolving isolated regions can be used to provide gradual decoder refresh. A new evolving isolated region is established in the picture at the random access point, and the macroblocks in the isolated region are intra-coded. The shape, size, and location of the isolated region evolve from picture to picture. An isolated region can be inter-predicted from the corresponding isolated region in earlier pictures in the gradual decoder refresh period. When an isolated region covers the whole picture area, a picture that is correct in content is obtained when decoding starts from the random access point.

Decoding of the leftover regions between a random access point and the corresponding recovery point is not necessary when decoding starts from the random access point and GDR is implemented using isolated regions. This can be indicated in the random access point SEI message of the JVT coding standard. Decoders can omit the decoding of the signaled leftover regions and save computations. A streaming server can implement the seek operation in on-demand streaming so that the coded leftover regions between the desired random access point and the corresponding recovery point are not transmitted. Servers can use the saved bitrate for forward error control or faster startup compared to conventional intra-slice-based GDR, for example.

The JVT codec includes a deblocking loop filter. Loop filtering is applied to each 4x4 block boundary, but loop filtering can be turned off at slice boundaries. If loop filtering is turned off at slice boundaries, perfect reconstructed pictures can be achieved when performing gradual random access. Otherwise, reconstructed pictures are approximately correct in content even after the recovery point. However, the mismatch is likely to be unperceivable. The random access point SEI message includes a flag to indicate if pictures after and at the recovery point are exactly or approximately correct in content.

5. SIMULATIONS

Two sets of simulations were conducted to access the coding efficiency and the error resilience property of the proposed GDR using isolated regions techniques.

Coding efficiency simulations. GDR based on isolated regions was compared to periodic IDR picture coding at a 1-second random access period. Error-free application environment, such as local storage, was assumed, and therefore the coding options yielding the best coding efficiency were selected. The simulations abided by the coding efficiency simulation common conditions specified by ITU-T Video Coding Experts Group [8]. A number of QCIF and CIF sequences were coded, and the average bitrate loss of GDR compared to periodic IDR was between 11 and 17%. Additional results can be obtained from [3].

Error resiliency simulations. The error resiliency performance of GDR was compared with the periodic IDR picture coding. The target was to simulate multicast streaming where random access points allow new receivers to start decoding. Random access period of about 1 second was used. The coded bitstreams were decoded after packet loss simulation under loss rates of 0, 3, 5, 10 and 20%. The results are presented in Fig. 2. It can be seen that GDR performs consistently better compared to periodic IDR in all tested loss rates above 0. Moreover, the PSNR difference between the cases grows as a function of loss rate.

6. CONCLUSIONS

Intra-coded pictures or progressively located intra-coded slices have been conventionally used to provide random access in coded video streams. This paper proposed a novel gradual decoder refresh method that is based on the isolated regions technique. The proposed method is especially suited for regularly occurring random access points that can be used to enable tuning in to a broadcast, for example. Furthermore, the paper described how the proposed method is included in the JVT coding standard and explained how the proposed method is more flexible and suits packet-based transmission better compared to progressively located intra-coded slices. The performed

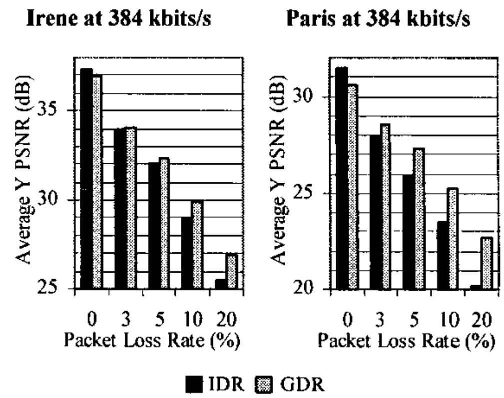

Fig. 2. Comparison of periodic IDR and GDR in terms of average luminance PSNR at different packet loss rates.

simulations revealed that the proposed method outperforms intra-picture-based random access points in error-prone network conditions.

7. REFERENCES

[1] T. Wiegand (ed.), "Study of final committee draft of joint video specification (ITU-TRec. H.264 | ISO/IEC 14496-10 AVC)," Dec. 2002. Available ftp://ftp.imtc-files.org/jvt-experts/2002_12_Awaji/JVT-F100d1ncm.zip.

[2] L. Wang, A. Luthra, and B. Eifrig, "Rate control for MPEG transcoders," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 2, pp. 222-234, Feb. 2001.

[3] Y.-K. Wang and M.M. Hannuksela, "Gradual decoder refresh using isolated regions," Joint Video Team document JVT-C074, May 2002. Available ftp://ftp.imtc-files.org/jvt-experts/2002_05_Fairfax/JVT-C074.doc.

[4] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proc. IEEE*, vol. 86, no. 5, pp. 974-997, May 1998.

[5] Y. Wang, S. Wenger, J. Wen and A.K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, vol. 17, no. 4, pp.61-82, July 2000.

[6] Y.-K. Wang and M.M. Hannuksela, "Error-robust video coding using isolated regions," Joint Video Team document JVT-C073, May 2002. Available ftp://ftp.imtc-files.org/jvt-experts/2002_05_Fairfax/JVT-C073.doc.

[7] P. Frossard and O. Verscheure, "AMISP: a complete content-based MPEG-2 error-resilient scheme," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 9, pp. 989-998, Sep. 2001.

[8] G. Sullivan and G. Bjontegaard, "Recommended simulation common conditions for H.26L coding efficiency experiments on low-resolution progressive-scan source material," ITU-T Video Coding Experts Group document VCEG-N81, Sep. 2001. Available ftp://standard.pictel.com/video-site/0109_San/VCEG-N81.doc.