Error Resilient Schemes for Region of Interest based Video Coding

[1]Mr P Subramanian, [2]Ms R Indumathi,

[1]*Associate Professor, ECE Department, Aarupadai Veedu Institute of Technology, Vinayka Missions University*

[2] *AssistantProfessor, ECE Department, Aarupadai Veedu Institute of Technology, Vinayaka Missions University*

Abstract:

In video applications, not all regions of a particular frame are of equal importance. There are one or more regions in a video frame that have higher importance than the rest of the frame. In medical videos, only part of the frame is of high diagnostic importance. The medically important region can be coded with higher bit allocation. This paper reviews the current error resilient, region of interest video coding schemes, which improve the quality of region of interest in video coding. Delivery of video is still a relevant problem today due to the numerous challenges in preserving information integrity in error prone environment such as the Internet and wireless networks. The effect of data loss in the transported compressed bit stream may lead to objectionable visual distortion at the decoder. There has been a lot of research to remedy or minimize the effect of channel errors on compressed video streams. This paper reviews the current schemes for making the region of interest based video coding more error resilient.

Keywords:
Error resilience, Bit rate, Region og Interest, Video Coding

I. Introduction:

The region of interest (ROI) video coding is an efficient scheme to enhance the quality of relatively important areas in a video frame [1]**,** such applications include video telephony and video conferencing. In such video applications, there exist one or more regions in one video frame have higher importance than rest of the frame. In the literature, ROI coding has been discussed as a tool of improving the perceptual quality of more important areas in a video sequence [2]-[9] and as an error resilience tool [10] [11]. Former is widely used in low bit rate conversational video communication applications since it provides a way of utilizing the available bit rate to maximize the effectiveness of the visual communication. This bit rate is often insufficient to retain more important visual clues in conversational applications such as facial expressions. However, low quality background is not unacceptable for the human visual system since it pays less attention to the background. Therefore, more bits can be allocated to the foreground without distracting the overall quality of the video. In this way, the foreground quality can be boosted. Chai el al proposed techniques called Maximum Bit Transfer (MBT) and Joint Bit Assignment (JBA) [3]. Former technique assigns the worst possible quantizer to the background and the foreground quantizer is adjusted accordingly such that the saved bits are consumed in the foreground. However, this technique does not take into account the fact that the worst quality on the background is not always desirable. Visual sensitivity of the human visual system has also been considered to determine the foreground and background quantization parameters [4]. Senguptha et al [5] has presented a technique in which the bit-allocation is based on a target quality for the foreground. This technique allows only perceptually tolerable graceful quality degradation for the background. Apart from the above algorithms, there are some algorithms developed for wavelet transform based coding techniques. Fukuma et al [6] [7] has proposed to use two wavelet filter sets with different tap lengths for ROI and leftover region. Advantage of such techniques is that it is not necessary to do any modification in the codec. Furthermore, the ROI boundary can be defined at a greater resolution, up to pixel level, unlike in the first approach, where the boundary must be defined in terms of macroblock boundaries. The technique proposed by Chen et al. [8] applies a low-pass filter to the background areas. Since they use only one filter over the entire background, it can cause boundary effects at the ROI border. Karlsson etal [9] use variable Gaussian filters controlled by a quality map indicating the distance to the ROI border. This approach is capable of achieving smoother ROI-to-leftover (and vise versa) transition. However, since these techniques are applied before the actual coding is taken place, they do not provide any help to the rate controller to decide how to utilize the available bit rate. In terms of improving error resilience, Jerbi et al [10] [11] have proposed to add redundant representation of the ROI. They have achieved this by applying a non-linear transform to duplicates macroblocks inside the ROI. The advantage of this technique is that it can be applied in a pre-processing stage and post-processing stages without modifying the source codecs. In the remaining sections we discuss in detail the most recent schemes namely flexible macroblock ordering, unequal error protection scheme, unequal error protection with forward error correction and non linear transformation.

II. Flexible Macroblock Ordering

Almost all the video coding standards follows a block based coding structure. Under this approach, the smallest block of picture to which the coding algorithms applied is called macroblocks. For example, MPEG and H.264 standards support 16x16 macroblocks. In order to add synchronization points to the resulting bit stream, a number of macroblocks are grouped into an entity called a slice. The slice may contain one or more macroblocks up to total number of macroblocks in a picture. Conventionally, video coding standards support encoding macroblocks only in raster order, keeping neighboring macroblocks next to each other in the encoded bit stream. However, with the introduction of FMO [12] in H.264 video coding standard, the order of assigning the macroblocks into slices has been liberalized. They can be grouped into entities known as slice groups freely. Subsequently, these entities are coded into separate Network Abstraction Layer (NAL) units [13] making them totally independent from others. Decoder identifies the particular macroblock assignment through a macroblock assignment map (MBA map) defined in the Picture Parameter Set (PPS) NAL unit associated with the picture. This map identifies which macroblock goes to which slice. There are six possible types of MBA maps. FMO type 0 macroblocks from each slice group are interleaved to fill the frame. Type 1 uses a function which is known to both the encoder and the decoder to disperse macroblocks over the picture. This is also known as scattered slices. FMO type 2 identifies certain rectangular areas of the picture and macroblocks from each area is assigned to different slice groups. FMO type 3, 4 and 5 are dynamic assignments which grow or shrink in a cyclic manner. Finally, the FMO type 6 is the most liberal type which allows fully random allocations. The main objective of the FMO is to improve the error resilience. For example, FMO type 0 and 6 can be used to arrange every other macroblock into separate slice groups like a check-board is formed. Macroblocks marked in black are coded into one slice group while the rest is coded into a separate one. Since each slice group is an independent entity loss of once slice group does not affect the decodability of the other. This arrangement improves the error concealment accuracy when one of these NAL units is corrupted or lost during the transmission. FMO Type 2 has been defined to handle ROI coding. The ROI can be defined as a rectangular region using this technique. However there would be some macroblocks which are not in the ROI are also included into the ROI. Therefore, part of the resource allocated to the ROI would be wasted. Therefore, this technique is not the best way to code an ROI. This scheme uses Type 6 FMO since the ROI can be defined more accurately. Flexible Macroblock Ordering (FMO) in H.264 provides an ideal framework to isolate certain areas of a video frame and encode them as separate entities. It has been proposed to use the FMO Type 6 to code the relatively important areas of a video frame (foreground) into one NAL unit, and the less important areas (background) into another NAL unit. One of these NAL units carries the background, and the other contains macroblocks from the foreground. Object locating and tracking techniques such as face tracking can be used to identify and track the movements of the foreground [2] [14]. Based on the boundary of the ROI, a set of FMO maps is defined to identify the ROI and to follow the temporal movements. When the object boundary changes significantly, a new picture parameter set (PPS) NAL unit containing the new FMO map is generated. In order to improve the quality over the ROI, the rate controller is modified since the existing rate controller treats all the slices equally. The leftover slice group of a picture is always coded first. The algorithm increases the QP determined by the native rate controller when the macroblock belongs to leftover region. Therefore, these areas are coded at a lower quality and consume fewer bits. Saved bits are now allocated for the second slice, which contains macroblocks from the ROI. Since more bits have been allocated for this slice, resulting quantization parameter is smaller. Therefore the quality over the ROI improves.

III. Unequal Error protection (UEP) Scheme for ROI.

This technique extends ROI video coding idea further by adding more error resilience into the foreground and applying unequal error protection (UEP) in which foreground is protected over the background. The video frame is encoded into three separate slices. First slice carries the background, and other two slices contain alternative macroblock of the foreground forming a check board pattern as shown in Figure 1. Figure 2 illustrates the generalized block diagram for the proposed UEP Scheme based codec. Since H.264 network abstraction layer (NAL) encodes each slice into separate data packet, single frame encodes into three different packets. Separating alternative macroblocks into separate slices improves the error robustness and helps better error concealment in case of bit error(s) or loss of packet(s). Foreground packets are protected with a stronger error correction code than background packets. This helps to save some channel capacity for encoding the foreground at a better quality. Experimental results show that this technique improves the objective quality of the foreground by more than 1 dB compared to the 1/3 EEP and up to 30 dB improvement over non-check board patterned ROI coding with the similar protection when SNR is in between 5dB and 10dB

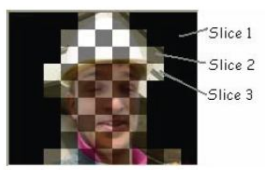

Figure 1. Slice assignment

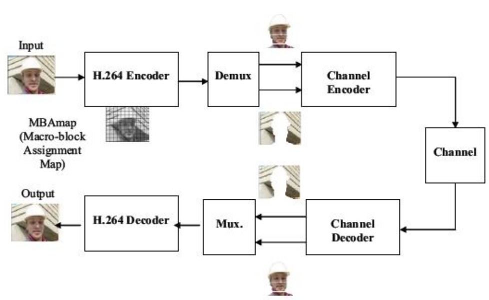

Figure 2. Block Diagram of UEP scheme based codec

IV. Combining UEP with Forward Error Correction.

This scheme incorporates an object mask generation using Chesnokov Yuriy's object tracking algorithm published in CodeProject [15]. The way of Yuriy's object tracking is to compare a pre-defined background with incoming captured images and draw each object's boundary (called blobs) based on their motion vectors and RGB differences. According to detected object mask, macro blocks can be divided into two separate regions: ROI region and Non-ROI region. Since motion vectors represent the motion dynamics within a frame, we further classify the macro blocks within the ROI region into two motion areas based on their motion vectors, one corresponding to large movement areas (or moving areas) and the other corresponding to still or small movement areas (or static areas). Now we can get three slice groups which have different priority after using FMO to group these regions to separate slice groups.

1) SG0: Non-ROI region with low priority.
2) SG1: Static areas in ROI region with medium priority and
3) SG2: Moving areas in ROI region with high priority.

When macro blocks in SG1 lost during transmission, it is easy to recover those by copy the macro blocks in the same position of previous frame since the movement in SG1 is very small even zero, but for macro blocks in SG2 loss, the recovery seems not so easy since in a large movement area it is hard to get an accurate MV for the lost macro blocks (the MV of lost MB is also missing along with the MB loss). Hence SG2 has been accorded the highest priority. In this scheme it is assumed that neighboring frames differ not significantly and hence the object mask and MVs of the previous frames can be used to determine the MB-to-Slice Grouping mapping in the next frame. For the first encode frame, we use a pre-defined rectangular windows as the center for Slice Group mapping, which is a reasonable approach since for most deliberately capture video, the ROI region is almost located at the center of a frame. The Slice Group mapping for the other frames will be updated. Since UEP transmission can allocate different error coding rate for different priority data [16], it is widely used in communication system. Three different priority Slice Groups are generated in our scheme, an UEP scheme should be adopted in video transmission. The high, medium and low protection orders FEC used in the simulation work are rate-compatible punctured convolution codes [17] encoder of rate 1/3, 1/2 and 2/3, respectively. Figure 3 shows the framework of UEP based on MB-to-Slice Groups Ordering. In RS encoder, we give the high priority data (SG2 packet) for more redundancy (lower code rate) and the low priority data (SG0 packet) for less redundancy (higher code rate). Experimental results have shown that this scheme not only improves the ROI-PSNR by 0.87dB in average but also achieves a much better subjective quality of the ROI region compared to the other UEP transmission schemes based ROI video coding under the same packet loss rate.

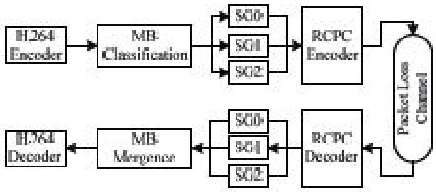

Figure 3. UEP based on MB-to-SG ordering

V Non Linear Transformation

Video coding standards that exist to date, exploit temporal redundancies using inter-frame prediction coding. Inter-frame coding is based on predicting the current source frame using the previously coded frame, and coding only the prediction error between the source frame and the motion-compensated predicted frame. Therefore, we not only need to consider coefficient loss as in the case for images or I frames, but also motion vector loss which usually has more devastating impact on the reproduced video quality. It can be assumed that the ROI does not change in the entire video sequence. This assumption is valid in video conferencing applications with head and shoulder type of content. This method, however, can be adapted to a more general class of video sequences where the ROI can be tracked from frame to frame. Figure 4 shows the proposed pre- and post-processing blocks in a typical video communication system. Since these blocks are outside the encoder and decoder functions, they can be widely applicable to all video coding standards such as H.263, H.264, MPEG-1,2,4. As shown in Figure 4, each frame $f_k(x,y)$ at time k undergoes a nonlinear transformation to produce a frame $g_k(x,yg)= f_k(x,y)$ where $y_g=T(y)$ and T is a non-linear transform.

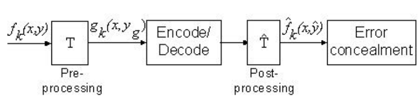

Figure 4. Overview of the system

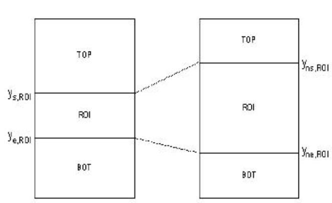

Figure 5. Frame $f_k(x,y)$ and the transformed frame $g_k(x,y_g)$

Figure 5 shows a frame before and after non-linear transform. In this figure $Y_{s,ROI}$ and $Y_{e,ROI}$ are boundaries of ROI that are transformed into $Y_{ns,ROI}$ and $Y_{ne,ROI}$ in the non-linear transformed frame. $Y_{s,ROI}$ and $Y_{e,ROI}$ which are determined by the user, are then adjusted so that the ROI boundaries would coincide with a macroblock boundary. The motivation behind the proposed non-linear transform is the fact that video coding is performed at a macroblock level. Indeed, the building blocks of a video codec such as inter-prediction compensation, intra prediction, etc…are often done one macroblock at a time. This suggests that if we duplicate a macroblock in the source frame, thereby generating a macroblock pair, then the corresponding reconstructed macroblock pair resulting from encoding and decoding will also look very similar to each other. Therefore, to make a bit stream more resilient to channel errors, it makes sense to duplicate data for each row of macro blocks. Referring back to Figure 1, the post-processing block, which occurs after decoding a frame, consists of an error recovery step and the inverse transform step. In the case where either coefficients or motion vectors for a particular macroblock are lost (a bad macroblock), and if its corresponding macroblock in the pair was received successfully (good macroblock), then the error recovery step will simply use the reconstructed data for the good macroblock to remedy the loss in the bad macroblock. In the case where both macroblocks in the pair undergo either coefficients or motion vectors loss, the pair is left to the concealment block (see Figure 1). For concealment, we simply use the reconstructed macroblock pointed to by the motion vector in the case where only coefficients loss occur to conceal the particular macroblock. When the motion vectors are lost, we use the median prediction among the motion vectors associated with the neighboring macroblocks to perform the concealment. This method yields an acceptable quality for the ROI within a head and shoulder type of video sequence even in the presence of packet loss. This is achieved by using a non-linear transform that increases the redundancy of macroblock rows within the ROI by generating macroblock pairs. This increased redundancy for the pixels within ROI yields a high quality reconstruction of ROI even with missing data.

References :

[1] A. P. Bradley, "Can region of interest coding improve overall perceived image quality?" in Proc. APRS Workshop on Digital Image Computing, Feb. 2003, pp. 41–44.

[2] A.Elefiheriadis and A.Jacquin, "Automated face location detection for model-assisted rate control in H.261 compatible coding of video", Signal Processing: Image Communication, Vol.3, pp.435-455, 1995.

[3] D. Chai, K. N. Ngan and A. Bouzerdoumu, "Foreground/background bit allocation for region-of-interest coding," in Proceedings of IEEE International Conference on Image Processing, Vancouver, Canada, vol. II, pp. 923-926, Sep. 2000.

[4]S.Daly, K.Matthews and J.Ribas-Corbera "Face based visually optimized image sequence coding," in Proceedings of International Conference on Image Processing, Chicago, USA, pp.443-447, Oct 1998

[5] Sengupta, S., Gupta, S.K., Hannah, J.M., "Perceptually motivated bit-allocation for H.264 encoded video sequences," in Proceedings of International Conference on Image Processing, Barcelona, Spain, Vol. 3, pp 797-800, Sept. 2003.

[6] S. Fukuma, S. Ikuta, M. Ito, S. Nishimuru, M. Nuwate, "An ROI Image Coding Based on Switching Wavelet Transform," Proceedings of the 2003 International Symposium on Circuits and Systems, Vol 2, pp. 420-423, May 2003.

[7] S. Fukuma, S. Ikuta, M. Ito, S. Nishimuru, M. Nuwate, "An ROI Image Coding Based on Switching Wavelet Transform," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E88-A No.7 pp.1995-2006, Jul. 2005.

[8] Mei-Juan Chen, Ming-Chieh Chi, Ching-Ting Hsu and Jeng-Wei Chen, "ROI Video Coding Based on H.263+ with Robust Skin- Color Detection Technique," IEEE Transactions on Consumer Electronics, Vol.49, No.3, pp.724-730, Aug. 2003.

[9] L. S. Karlsson and M. Sjostrom,"Improved ROI Video Coding using Variable Gaussian Pre-Filters and Variance in Intensity," in Proceedings of IEEE International Conference on Image Processing (ICIP) 2005, Genova, Italy, Vol. II, pp. 313-316, Sep 2005.

[10] A. Jerbi, Q.Wang, and S. Shirani, "Error-resilient ROI coding using pre and post- processing for video sequences," in Proceedings of IEEE International. Symposium on Circuits and Systems, pp. III- 757-60, May 2004.

[11] A. Jerbi, Q.Wang, and S. Shirani, "Error-Resilient Region-of- Interest Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 9, pp. 1175-1181, Sep. 2005.

[12] Michael Horowitz, Stephan Wenger, "Flexible Macroblock Ordering (FMO) 101," Input Document to JVT: JVT-D063, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 4th Meeting: Klagenfurt, Austria, Jul., 2002.

[13] T. Stockhammer, M. Hannuksela, S. Wenger, "H.26L/JVT Coding Network Abstraction Layer and IP-based Applications", IEEE International Conference on Image Processing Rochester, NY, pp 485-488, Sep. 2002.

[14] Y.-O. Kim, J. Paik, H. Jingu, A. Koschan, B. Abidi, M. Abidi, "Automatic face region tracking for highly accurate face recognition in unconstrained environments," in Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance, Miami, USA, pp. 29- 36, Jul. 2003.

[15] [On-line] http://www.codeproject.com/ KB /audio-video/ project_track er.aspx.

[16] Wang. Y, etc. "Rate-distortion optimized layered coding with unequal error protection for robust internet video", IEEE Transactions on circuit and systems for video tech, March 2001.

[17] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," IEEE Trans. Commun., vol. COM-36, pp. 389-400, Apr. 1988.

[18] H Kodikara Arachchi et al "Unequal Error protection technique for ROI based H.264 Video coding", IEEE CCECE, May 2006

[19] Xiaolong Wang, Kun Tang, Huijuan Cui, " A novel unequal error protection scheme for ROI based video coding in H.264 /AVC", Second International Conference on Computer Modelling and Simulation, 2010

[20] H Kodikara Arachchi et al, " Region of Interest Video Coding with flexible macroblock ordering", ICIIS Srilanka, August 2006,