A Content-based Robust Mode Decision Scheme for Internet Videophone Applications*

Jin Xu[1,2], Limin Sun[1] and Zhimei Wu[1]
[1]Institute of Software, Chinese Academy of Sciences, Beijing 100080, China
[2]Graduate University of Chinese Academy of Sciences, Beijing 100039, China
xujin@iscas.ac.cn

Intra-coding macroblocks (MBs) is a fairly efficient way to make the compressed video bitstream more resilient to channel error because the decoding of an intra MB does not need the information from previous frames. However, too many intra-coded MBs will significantly degrade the compression performance. Therefore, a good mode selection between intra-mode and inter-mode should be in place to enhance the robustness of the video communications system to channel error. Since Rate-Distortion optimized (RDO) video coding framework can provide an efficient means for coding mode selection in an error-free environment, recent researches propose to extent this RDO framework in error-prone environment to intelligently decide which MBs should be intra-coded to better combat the transmission error. However, since all these conventional error resilient RDO mode decision schemes generally use objective distortion measurement metrics such as mean square error (MSE) to calculate the overall distortion (i.e., the source coding distortion plus channel distortion) and the same amount of distortion measured in MSE at different position within a scene does not mean the same perceived quality to the human eyes, therefore, the conventional error resilient RDO mode decision schemes can not guarantee to produce an intra-refreshing pattern according with the characteristic of human visual system (HVS). Another intuitive mode selection strategy for videophone applications is solely intra-coding the MBs belonging to face regions. Although such an intuitive mode decision scheme can keep important information associated with face regions free from channel errors, it degrade the compression performance significantly especially when the face regions take up a big part of a video frame and have little motion.

In this paper, a content-based error resilient RDO mode decision scheme is proposed for Internet videophone applications. First, since MB is the basic encoding unit in most video coding standards, the visually important regions within a head-and-shoulder video sequence are determined by using a simple yet efficient MB-based skin segmentation technique. Then, a skin-region-weighted MSE is used to calculate each MB's potential error-propagated distortion on a block basis (e.g. 4x4). For a block belonging to the skin regions, its potential error-propagated distortion will be scaled by a larger weighting factor; otherwise, it will be scaled by another smaller weighting factor. Therefore, the potential error-propagated distortion of a block belonging to visually important regions will be increased faster than that of visually unimportant regions, and then the visually important MB will have a bigger chance to be intra-coded in the RDO mode selection process. The experimental results based on typical head-and-shoulder video sequences under common Internet channel condition show that, comparing with conventional error resilient RDO mode decision schemes, the proposed mode decision scheme improves the average luminance PSNR of visually important foreground area consistently regardless of the packet loss rate. This improvement is gained at the expense of degraded background image quality because the macroblocks belonging to background area are less frequently intra-coded. However, because human eyes generally pay more attention to the foreground objects in a videophone application, errors in background area are far less noticeable than errors in the foreground area, and therefore the overall subjective quality is improved.

* This work is supported by the National Natural Science Foundation of China under grant No. 60673178.

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE