J. Vis. Commun. Image R. 20 (2009) 491–503

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Adapting quantization offset in multiple description coding for error resilient video transmission ☆

Viswesh Parameswaran, Avin Kannur, Baoxin Li *

Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85281, USA

ARTICLE INFO

Article history:
Received 21 August 2008
Accepted 4 July 2009
Available online 12 July 2009

Keywords:
Multiple description coding (MDC)
Video streaming
H.264/AVC encoder
Error resilient video transmission
Wireless ad hoc networks
Video quality
Video over wireless
Quantization offset

ABSTRACT

Multiple description coding (MDC) provides an excellent error resilient approach for transmitting video over wireless ad hoc networks. In this paper, we propose an improvement to this scheme by jointly selecting the quantization offsets in different paths to achieve the best overall video quality at the decoder. The statistical distribution of the transform coefficients in the encoded video sequence is parameterized using a Laplacian model. The optimal offset values are computed by solving a multi-level optimization problem based on the statistics of the transform coefficients and the individual path failure probabilities. In order to reduce the computational complexity, the encoding modes for the motion vectors and transform coefficients are collected in the first step. In the second step, the model parameter is calculated from the transform coefficients and the offset search is performed when the model parameter between frames deviates beyond the pre-set threshold. In the final step, the stored modes and the respective offset values are directly used for encoding the two bit streams. The simulation results using H.264/AVC system confirm the advantages of the proposed approach under different packet loss conditions.

1. Introduction

The past decade has witnessed rapid growth in the development and deployment of computationally intensive multimedia applications on wireless devices. This can be attributed to the following reasons: – (1) the improvements in semiconductor technologies resulting in the development of 45 nm integrated chip fabrication process, (2) the developments in wireless technologies enabling information transfer at speeds ranging 2 Mb/s in the case of 3G networks, and (3) the evolution of video compression standards enabling high quality video transmissions at low bit rates. Some of the cell phones available in market are already equipped with digital video streaming and conferencing functionalities along with the fundamental voice communication feature. The anticipated growth in mobile social networking will further increase the utility of multimedia transfer over cell phones. The typical video applications running on these devices can be categorized on the basis of their underlying network, delay requirements and the desired video quality. The network used for video transfer could be fixed wire line, fixed wireless, ad hoc wireless or a combination of wired and wireless networks. The delay parameters that need to be considered for system design are the average end-to-end delay (known as *latency*) and the delay variation (known as *jitter*). The video quality can be specified in terms of the required spatial and temporal resolutions.

For instance, consider the case of one-way video streaming over internet. In this application, the video is compressed offline and is stored in a streaming server. The user connects to the server using a suitable protocol such as real-time streaming protocol (RTSP). Since the video is encoded offline, there is no delay constraint at the encoder. The decoder stores the received frames in a buffer and starts playing them after an initial play-out delay. In this case the latency can be fairly large, but jitter should be limited so that the video can be played out smoothly. The maximum jitter that can be tolerated is determined by the size of the playout buffer. The desired video quality in the case of this application is determined by the end user terminal and the speed of the linking network.

Another application scenario that can be envisioned is interactive video conferencing over cell phones. In this case we need to capture and deliver video in real time. In order for the two-way communication to be effective, we need to impose stringent delay constraints both at the encoder and the decoder. The system has to be designed in such a way so as to minimize latency and jitter. The buffers at the encoder and decoder should be of limited size to minimize latency. The underlying network in this case is wireless and hence the application is constrained by limited band-width

☆ This research was supported in part by ARO Grant W911NF-06-1-0354. The information reported here does not reflect the position or the policy of the federal government.

* Corresponding author.
E-mail addresses: vparames@asu.edu (V. Parameswaran), akannur@asu.edu (A. Kannur), baoxin.li@asu.edu (B. Li).

1047-3203/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2009.07.003

and high error rates. The error characteristics of the channel fluctuate because of the multipath fading effects. The video quality requirements are relaxed in this case since we can tolerate some degradation in terms of spatial and temporal resolution.

Similar constraints are applicable in case of mobile imaging sensors as well. Recently there has been a lot of focus in developing miniature low-power imaging sensors by companies such as SONY, CANON, Cypress semiconductors etc. These sensors are finding increasing number of applications in military and commercial environments. For example, in a typical military battle field scenario, the imaging sensors can be mounted on a soldier's helmet and the captured video could be relayed to the base station at a remote location. Alternatively, the sensors could be mounted on unmanned ground/air vehicles (UGV/UAVs) and the captured video could be transmitted to the soldiers at a different location. This shared vision approach helps in minimizing the risk of injury and enhances the tactical options available to the soldiers.

The basic problem in all the above applications is video transmission over wireless networks. The wireless channel is inherently error prone because of omnipresent noise and multi-path fading. The use of forward error correcting codes (FEC) is a popular way of providing error resilience in these conditions as demonstrated in articles [2,9,15,25,29] etc. The channel codes work well as long as the packet drop probability is low. In case of burst errors or path outages, the use of FEC alone is not sufficient to provide adequate error resilience. This is especially true in case of low-power mobile imaging sensors, where the transported video stream is very likely to use compression techniques such as H.264 thereby having inter frame dependencies which cannot be handled well using FEC. FEC based on popular channel codes do not seem to provide the best performance at high packet drop rates. In this case, video from the source goes over multiple hops before reaching the destination. Since each of these nodes is mobile, any node can move out of the transmission range resulting in path outage. In some cases, the best available path might change over a period of time and switching from one path to another will become necessary. In this scenario, it is advisable to transmit multiple correlated descriptions of the same video source over disjoint paths. Such a coding scheme, termed as multiple description coding (MDC) [5,17], provides error resilience through path redundancy. A good comparison between MDC and single description coding with FEC is presented in [3]. In MDC, the individual descriptions can be decoded independently of each other. In case we receive both descriptions the information from them can be combined together to generate a better quality video at the decoder. This is in contrast to a layered coding system, where the base layer and the enhancement layer information are transmitted over different paths. In such a system, the enhancement layer information cannot be decoded independently from the base layer. Typically the layered coding is suited for scenarios where one path is highly reliable compared with the other path. If both paths are equally good or equally worse, it is better to use MDC for achieving error resilience. The history and development of MDC for audio and still image applications was covered in detail by Goyal in [5].

Some of the popular approaches for multiple description video coding were explored by Brian in [7]. The MDC scheme for encoded video transmission gets complicated because of drift error propagation between frames. The latest video coding standard H.264/AVC, [30,18,28], employs both spatial intra-prediction as well as motion compensated inter-prediction for achieving high compression ratios. In these cases the loss of one packet can cause unpredictable impact in the current frame and future dependent frames. In order to better understand the drift effects, we will take two popular approaches for MDC – temporal sub-sampling (MDTS) and repetitive coding (MDRC) and compare them in terms of their coding efficiency and error resilience. In MDTS, the original video sequence is partitioned temporally into even and odd frames. These two descriptions are encoded independently and transmitted over two different paths. In case both descriptions are received at the decoder, the decoder interleaves the received frames and reconstructs the original video. The problem happens when only one description is received. In this case we cannot use the information from the other description to reconstruct the lost frame. The only option available is to do error concealment as in case of single description decoder and this results in propagation of errors in future frames. The coding efficiency is also reduced in this scheme because of poor temporal correlation. In MDRC, identical descriptions are transmitted over both paths. This scheme is highly error resilient since the loss of one description can be compensated directly with information from the other path. In case both descriptions are received, the information from one path is discarded. This leads to a overall loss of performance especially at low packet drop rates. The MDC scheme, in general, is suited for high packet drop networks or networks with path outages. In addition to this highly efficient error concealment techniques, similar to the ones covered by Wang in [24,26], can be implemented at the decoder with MDRC scheme since both descriptions contain information about the same frame.

Hence in this paper we propose a modified RC based MDC scheme that relies on the optimal selection of the quantization offsets for the transform coefficients to achieve the best overall video quality at the decoder [27]. The design of multiple description scalar quantizers (MDSQ) was first done by Vaishampayan in [21–23]. In this scheme, two separate indices are generated for each quantization level in each path. Such a scheme can be implemented in practice by shifting the quantization levels by half in one of the paths. If both descriptions are received at the decoder, termed as the central description, a finer quantized video can be generated. But in case only one side description is received the quality deteriorates because of the additional offset. An improvement of the scheme was proposed by Lee in [12], where the spatial and motion information is used to reconstruct the video in case only one description is received. The scheme reported above shows good performance for intra-predicted MBs, but does not show significant gains in case of inter-predicted MBs. Jafarkhani and Tarokh describe a MDC scheme based on trellis-coded quantization in [8].

In our approach, we address the above problem by optimally selecting the quantization offsets depending on both the statistics of the input data source as well as their respective path failure probabilities. This is done because the overall video quality at the decoder does not depend solely on the individual side description or the central description, but instead is a function of the side and central distortion weighed by the probabilities of receiving each of them. The side and the central distortion terms in turn depends upon the statistics of the input data source. We will do a thorough analysis on the variation of model parameter for luminance and chrominance transform coefficients. We will also demonstrate that the model parameter needs to be computed separately for the intra and the inter predicted frames and also for different transform domain block sizes. The motion vector and the transform coefficients prediction modes are collected in the first pass and are used for subsequent encoding of the two bit streams. The offset search is refined when the model parameter between frames varies beyond a pre-set threshold.

The rest of the paper is structured as follows. In Section 2 we will formulate the optimization problem to minimize the overall decoder side distortion based on the transform coefficients and path failure probabilities. Section 3 will cover the implementation details specific to H.264/AVC system. In Section 4 we will illustrate the block schematic of the proposed system. The simulation results under different packet loss conditions are presented in Section 5 with concluding remarks in Section 6.