Available online at www.sciencedirect.com
ScienceDirect
Signal Processing: *Image Communication* 22 (2007) 809–832
SIGNAL PROCESSING: IMAGE COMMUNICATION
www.elsevier.com/locate/image

Adaptive bitstream switching of scalable video ☆

Osama A. Lotfallah[a,1], Geert Van der Auwera[b], Martin Reisslein[b,*]

[a] *Johnson Controls Inc., Milwaukee, WI 53202, USA*
[b] *Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA*

Received 30 October 2006; received in revised form 8 June 2007; accepted 11 June 2007

Abstract

With scalable video coding that provides fine-granular quality degradation, such as fine granularity scalability (FGS) and progressive FGS (PFGS), or H.264 scalable video coding's (SVC) adaptive reference FGS (AR-FGS) coding, video can flexibly be streamed to receivers of heterogeneous bandwidths. However, the transmitted video is only efficiently encoded when the transmission bit rate is in the vicinity of the encoding bit rate. In this paper, we develop and evaluate a comprehensive suite of network-aware adaptive bitstream switching policies for point-to-point and point-to-multipoint streaming of fine granular scalable coded video to address this coding efficiency issue. Our approach stores a small number of encodings (versions) with different encoding bit rates for each video sequence and estimates the reconstructed quality using the motion activity levels of the underlying visual content (or, in general, any content descriptor(s) that highly correlate with the reconstructed quality). For unicast streaming, we then: (i) adaptively switch between the different encodings at the server, to improve the reconstructed video quality and (ii) adaptively drop packets during network congestion to ensure fairness between multiple unicast streams. For multicast streaming, we also adaptively switch between the different encodings to maximize the average video quality. Our adaptive bitstream switching policies consider the visual content descriptors as well as the network channel variability, while requiring only sample points from the rate-distortion curve of the video stream. From our extensive simulations with PFGS coding, we find that our adaptive unicast bitstream switching policy achieves on average a 0.8 dB improvement over the optimal non-adaptive streaming for a diverse 200-shot sequence from *Star Wars IV*. We have also verified our key findings with the latest scalable video coding standard, H.264 SVC.

Keywords: Adaptive streaming; Congestion control; Motion activity; Multicast; PFGS; Simulcast; SVC AR-FGS

1. Introduction

A key challenge of video delivery is to regulate the video transmission rate according to the network's and receiver's capabilities, i.e., to adapt the video transmission in a network- and client-aware manner. A potential approach for network- and client-aware video adaptation is the so-called simulcast or bitstream switching technique, which encodes

☆ A preliminary presentation of parts of this work appears in the Proceedings of the ACM Workshop on Advances in Peer-to-peer Multimedia Streaming 2005 [1].

*Corresponding author. Tel.: +1 480 965 8593; fax: +1 480 965 8325.
E-mail addresses: Osama.Lotfallah@jci.com (O.A. Lotfallah), Geert.Vanderauwera@asu.edu (G. Van der Auwera), reisslein@asu.edu (M. Reisslein).

[1] This work was conducted while O. Lotfallah was with Arizona State University, Tempe.

0923-5965/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2007.06.002

a given video at many different rates, i.e., into different versions, and then transmits the version with the highest rate that fits into the available network bandwidth. These techniques, while simple, have a number of significant drawbacks, such as the encoding of an impractically large number of versions (on the order of several tens of versions are required to adapt a Mbps video at the granularity of 100 Kbps). Also, there is no flexibility to scale down the bit rate/video quality of a stream during network transport, unless typically computationally demanding transcoding is performed at intermediate network nodes. Bitstream switching can also be applied over versions of different coding schemes, but an increase in the computational complexity of video decoding is inevitable [2–4]. Scalable (layered) video coding overcomes these drawbacks by encoding a video into a base layer and several enhancement layers. The base layer represents the basic video quality, which is typically transmitted with higher protection (often achieved with unequal error protection), and the enhancement layers, which are transmitted with lower protection, gradually improve the video quality [5–8]. A key limitation of layered video coding is that the video bit rate can only be adapted at the granularity of complete enhancement layers, whereby the number of layers is typically limited to a small number (at most 4–5 in practical encoders) resulting in rather coarse rate adaptation. Fine granularity layered coding techniques overcome this shortcoming by encoding the video into one base layer and one enhancement layer, whereby the enhancement layer bit rate can be finely adapted [6,9]. This flexibility in bit rate adaptation comes at the expense of relatively low compression efficiency. Progressive fine granularity scalability (PFGS) coding (also called two-loop fine granularity scalability (FGS) coding) overcomes this disadvantage and provides generally good compression efficiency as well as high flexibility in adapting the enhancement layer bit rate [10,11]. We verify our key PFGS findings with the H.264 adaptive reference FGS (AR-FGS) encoder [12] to assert that the same switching principles apply to state-of-the-art FGS schemes.

FGS coding makes it possible to encode the enhancement layer of the video with one (high, say 2 Mbps) bit rate and then to flexibly transmit the enhancement layer at any lower bit rate. However, there are compression inefficiencies: *for transmission at a low bit rate (say around 1 Mbps) the video could be coded (with the PFGS codec) much more efficiently by employing an enhancement layer encoding bit rate in the vicinity of the transmission bit rate. We quantify this efficiency gain achieved by selecting the encoding bit rate reasonably close to the actual transmission bit rate in Section 3.2 and demonstrate that it can reach on the order of 4 dB for some visual contents.* We also verify in Section 3.2, that for AR-FGS the quality gain can reach up to 1.8 dB for the same content. In summary, PFGS encoding—and, in general, any SNR scalable coding concept known today—is only optimized for the rate that is used during the encoding process [1,13]. Essentially all existing studies on PFGS video streaming have ignored the important aspect of the encoding bit rate selection and focus primarily on modeling the rate-distortion (R-D) characteristics of a given encoding or the optimal selection of the transmission bit rate based on the R-D model of one encoding [14,15].

Our main contribution in this paper is to develop and evaluate a comprehensive framework of network-aware adaptive bitstream switching policies for the streaming of scalable (layered) coded video. Our adaptive bitstream switching policies exploit video content features and cover common unicast and multicast streaming scenarios. While our framework applies generally to a wide range of scalable (layered) video coding schemes, to fix ideas, we focus on PFGS coding in our specific problem formulations. We optimally select the values of the encoding parameters, depending on visual content and network (channel) conditions so as to maximize the reconstructed video quality. Hence, we propose network-aware techniques for multimedia delivery using in-network processing application that utilizes the visual content descriptors, which can be stored at the video server for multimedia indexing as well as delivery adaptation. Specifically, in the context of PFGS streaming, we optimally adapt the enhancement layer encoding bit rate. To the best of our knowledge, this important encoding parameter setting problem has to date only been observed in [13] but no solution has been proposed, as detailed in Section 1.1. Our main approach is to pre-encode a given video with fine granularity scalable coding (such as PFGS coding) into a small number of versions with different encoding parameter values (e.g., different enhancement layer coding bit rates in the case of PFGS coding). In a sense, our approach combines simulcast with scalable coding. In contrast to simulcast, which requires many versions to adapt