Signal Processing: *Image Communication* 30 (2015) 89–106

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

A parallel H.264/SVC encoder for high definition video conferencing

Sergio Sanz-Rodríguez [a,*], Mauricio Álvarez-Mesa [a], Tobias Mayer [b], Thomas Schierl [c]

[a] *Embedded Systems Architecture, Technische Universität Berlin, Berlin, Germany*
[b] *Image Communication Group, Technische Universität Berlin, Berlin, Germany*
[c] *Multimedia Communications Group, Fraunhofer HHI, Berlin, Germany*

ARTICLE INFO

Article history:
Received 4 April 2014
Received in revised form
7 September 2014
Accepted 8 October 2014
Available online 22 October 2014

Keywords:
H.264/Scalable Video Coding (SVC)
Video conferencing
Rate control
High definition
Ultra-low delay
Parallel processing

ABSTRACT

In this paper we present a video encoder specially developed and configured for *high definition* (*HD*) video conferencing. This video encoder brings together the following three requirements: *H.264/Scalable Video Coding* (*SVC*), *parallel encoding on multicore platforms*, and *parallel-friendly rate control*. With the first requirement, a minimum quality of service to every end-user receiver over Internet Protocol networks is guaranteed. With the second one, real-time execution is accomplished and, for this purpose, slice-level parallelism, for the main encoding loop, and block-level parallelism, for the upsampling and interpolation filtering processes, are combined. With the third one, a proper HD video content delivery under certain bit rate and end-to-end delay constraints is ensured. The experimental results prove that the proposed H.264/SVC video encoder is able to operate in real time over a wide range of target bit rates at the expense of reasonable losses in rate-distortion efficiency due to the frame partitioning into slices.

1. Introduction

The increasing advances in video compression standards, network infrastructures as well as visual display technologies have made *high definition* (HD) video conferencing one of the most popular multimedia applications over Internet Protocol (IP) networks. Specifically, a video conferencing session involves point-to-point or multipoint real-time video and audio communication for multiple users that possibly are geographically spread, thus resulting in a challenge for video codec designers in order to provide real-time HD video content delivery with a minimum guaranteed *quality of service* (QoS). To this end, the following three key requirements are expected to be considered for a video coding system: an *H.264/Scalable Video Coding* (*SVC*)*-based approach*, a *parallel* (*multi-core*) *computing architecture*, and a *parallel-friendly rate control algorithm* (*RCA*). These requirements are described in the sequel:

- The scalable extension of the H.264/Advanced Video Coding (AVC) standard, named as H.264/SVC or simply SVC [1,2], is capable of delivering high-quality video content adapted to certain QoS imposed by either on-the-fly varying network conditions or the heterogeneity, in terms of display resolutions and computational capabilities, of end-user devices. The use of SVC involves the extraction of either one or a subset of sub-streams from a high-quality bit stream, so that these simpler sub-streams, bearing lower spatio-temporal resolutions or reduced quality versions of the original sequence, can be decoded by a given target receiver. For example, in a

* Corresponding author. Tel.: +49 30 31421 357; fax: +49 30 31422 943.
E-mail addresses: sergio.sanz@aes.tu-berlin.de (S. Sanz-Rodríguez), mauricio.alvarezmesa@tu-berlin.de (M. Álvarez-Mesa), tobias.mayer@hhi-extern.fraunhofer.de (T. Mayer), thomas.schierl@hhi.fraunhofer.de (T. Schierl).

http://dx.doi.org/10.1016/j.image.2014.10.003
0923-5965/

video conference session consisting of two target HD receivers, SVC could be used to generate a complete bit stream consisting of two dependency (spatial or quality) layers: a base layer that includes the sub-stream with low-quality compressed video, e.g., 720p@30 *frames per second* (fps), and an enhancement layer that includes additional information to deliver the high-quality version of video content, e.g., 1080p@30 fps. Thus, whereas for those low-quality receivers the enhancement layer is dropped to decode only the base layer, for the rest of target receivers the complete bit stream is delivered, unless the current network conditions are not suitable to transmit the whole bit stream and only the base layer must be decoded to get the best possible video quality for such conditions. Furthermore, unlike other well-known coding technologies, such as *simulcasting* and *transcoding*, SVC also provides the following benefits for video conferencing: (1) SVC is able to reduce the transmission bandwidth when compared to simulcasting, since the redundancies between the different video versions are actually exploited; and (2) due to the fact that the SVC bit stream itself contains all the video versions demanded by the application, no additional transcoding is required, thus reducing the end-to-end delay and, therefore, making the live session more natural.

- In order to accomplish real-time operation, the execution time of the encoder must be below the limits of the target frame rate, e.g., 33.33 ms per frame for 30 fps. For improving the time performance, typically real-time video encoders restrict the available encoding tools, with an acceptable loss in *rate-distortion* (R-D) efficiency, and use also platform specific optimizations such as *single instruction, multiple data* (SIMD) instructions [3]. The computational requirements of the encoder, however, exceed the capabilities of a single conventional processor, specially when processing HD content combined with a multilayer coding approach such as SVC. In addition to that, processor frequency is not increasing with every technology generation at the same rate as in the past; instead processor manufacturers are building systems with multiple processors (also called *cores*) per chip [4,5]. Then, in order to achieve real-time operation for multilayer HD coding, parallelization is necessary, and it must scale so that the performance improves with the growing number of cores per chip [6]. It should be noted that, when using SVC for video conferencing, the encoder must be able to process every access unit (defined as the union of all the representations of a picture at a given time instant) within the time limit of the target frame rate (e.g., the same 33.33 ms for 30 fps), and at the same time maintain a low end-to-end delay. Due to this, parallel techniques such as frame-level parallelism or *group of pictures* (GoP)-level parallelism that increases the throughput but do not reduce the frame latency are not well suited. Furthermore, parallelization techniques have to be applied not only to the single layer encoding scenarios where most of the execution time is spent in the main coding loop (motion estimation being the most complex part), but also to other functions in SVC such as upsampling filters for spatial scalability that can take an important part of the execution time. As a result, a parallelization strategy for SVC real-time encoding for video conferencing must be able to provide the required performance at the access unit level, while at the same time reduce the frame latency, and must take into account the additional processing steps using in multilayer applications.
- The variable bit rate nature of compressed video implies that an RCA must be embedded in the video encoder to avoid encoder buffer (and decoder buffer, which performs the complementary process) overflow and underflow, while providing as good as possible the quality consistency and R-D performance [7]. Furthermore, given that the ultra-low delay restriction in a video conferencing environment necessarily entails the use of very small buffer sizes, the RCA must also ensure a tight short-term *target bit rate* (TBR) adjustment. To achieve this, the *quantization parameter* (QP) of transform coefficients can be adjusted for every video segment, typically with size of *macroblock* (MB) in low-delay applications. For a proper selection of the QP value, the RCA should properly assign a bit budget for the current video segment considering the video complexity, the specified TBR as well as the *hypothetical reference decoder* (HRD) constraints [8] required to provide deliverable bit streams. It is also worth noticing that, when using slice-level parallelism in a video conferencing application, independent MB-level QP decisions within a picture must be conducted, so conventional RCAs are not longer valid unless a picture-level QP decision strategy is adopted at the expense of higher instantaneous bit rate variations (see Section 2.2). In short, an RCA for HD video conferencing should have the following two attributes: *low-complexity* and *parallel-friendly*. The former is recommended to facilitate real-time encoding, whereas the latter is required to provide accurate MB-level QP selection within a slice and, hence, strict buffer control.

In this paper we propose a complete video coding framework for HD video conferencing. Specifically, the SVC standard was used to guarantee a minimum QoS for every end-user receiver. In order to achieve real-time operation, a parallelization strategy that combines slice-level parallelism, for the main encoding loop, and block-level parallelism, for the upsampling and interpolation filters, was implemented. Furthermore, a novel low-complexity parallel-friendly RCA operating at MB level was embedded in the SVC encoder for a proper video content delivering. All these tools will be described in detail later on.

The paper is organized as follows. In Section 2 previous approaches related to parallelism for real-time video coding as well as the state of the art in RC for video conferencing are described. In Section 3 an overview of the SVC standard is given. In Section 4 the optimized SVC encoder is described in detail, emphasizing on those operations that were parallelized. In Section 5 a detailed description of the proposed MB-level RCA is given. In Section 6 the experimental setup is described and the results are reported and discussed. Finally, in Section 7 conclusions are drawn and future work is outlined.