J. Vis. Commun. Image R. 38 (2016) 378–385

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Multiple description video coding based on adaptive data reuse ☆

Meng Dong [a], Huanqiang Zeng [a,*], Jing Chen [a], Canhui Cai [a], Kai-Kuang Ma [b]

[a] *School of Information Science and Engineering, Huaqiao University, Xiamen, China*
[b] *School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore*

ARTICLE INFO

Article history:
Received 7 February 2015
Revised 23 November 2015
Accepted 17 March 2016
Available online 22 March 2016

Keywords:
Multiple description video coding
Adaptive data reuse
Error resilient
Inter prediction
Intra prediction

ABSTRACT

In this paper, an efficient *multiple description coding* (MDC) method, called the *adaptive data reuse MDC* (ADR-MDC), is proposed for robust video transmission. The proposed ADR-MDC first groups four sub-sequences, which are generated by performing spatial down-sample operation on each frame of input video sequence, into two descriptions. In each description, one sub-sequence is directly encoded and decoded by applying the H.264/AVC encoder and decoder, while the other is encoded and decoded at each *macroblock* (MB) based on the adaptive data reuse criterion, which are developed by making full use of the spatial and temporal correlations among the sub-sequences. Experimental results have shown that the proposed ADR-MDC scheme possesses higher error resilient ability and obtains better reconstructed video than the existing state-of-the-art MDC method.

1. Introduction

Video transmission over unreliable network is inevitably suffered from bit error and/or packet loss, which will significantly deteriorate the quality of the reconstructed video. For that, *multiple description coding* (MDC) is developed as an effective error-resilient lossy coding scheme for data transmission over error-prone channels [1,2]. In general, MDC encodes the original signal into multiple bit-streams. Each bit-stream is regarded as one description and each description can be independently decodable. Hence, these generated multiple descriptions can be individually transmitted through different channels to combat severe channel errors. If only one description is received, a baseline signal can be reconstructed. And the quality of the reconstructed signal at the decoder will be gradually increased with the number of the received descriptions.

Since the inception of the H.264/AVC [3] video coding standard, some multiple description video coding methods for robust video transmission have been proposed. Bernardini et al. [4] proposed a four-description coding scheme, called the *polyphase spatial subsampled* MDC (PSS-MDC), in which the input video sequence is first sub-sampled using the polyphase spatial decomposition scheme to generate four sub-sequences, followed by encoding them individually using the H.264/AVC encoder to produce four encoded descriptions. However, the error resilience gained in this method is at the heavy cost of coding efficiency and computational complexity. Based on the framework of PSS-MDC, Wei et al. [5] proposed an improved MDC scheme, called the *prediction-based spatial polyphase transform* MDC (PSPT-MDC), to improve the coding efficiency and mitigate the computational complexity. In this method, four sub-sequences produced by the polyphase down-sampling process were further grouped into two descriptions via the quincunx manner. To exploit the correlation between two sub-sequences within the same description, only one sub-sequence will be encoded by applying the H.264/AVC encoder (called the *directly encoded sub-sequence*), while the other sub-sequence (called the *indirectly encoded sub-sequence*) will be first predicted based on the directly encoded sub-sequence using a neighboring prediction algorithm, followed by encoding its resultant prediction errors. Since the neighboring prediction algorithm explores the spatial correlation of the neighboring sub-sequences (i.e., inter sub-sequence correlation), there is no need to compute, encode, and transmit *motion vectors* (MVs) and other relevant auxiliary data. Consequently, the coding efficiency of the PSPT-MDC [5] is much higher than that of the PSS-MDC [4]. It is worthwhile to mention that the framework of PSPT-MDC [5] has been successfully exploited for conducting multiple description image coding as presented in [6], together with the newly developed *adaptive redundancy control* (ARC) scheme that yields optimal tradeoff between coding efficiency and error resilience. Considering the redundant representation at the slice level inherited in the H.264/AVC, the *redundant slice-based* MDC (RS-MDC) was presented in [7]. If one description got lost, the redundancy at the slice level can be used to partially recover the missing data. To improve the error resilience and

---

☆ This paper has been recommended for acceptance by Yun He.

* Corresponding author.

E-mail addresses: mengyingyidai@163.com (M. Dong), zeng0043@hqu.edu.cn (H. Zeng), jingzi@hqu.edu.cn (J. Chen), chcai@hqu.edu.cn (C. Cai), ekkma@ntu.edu.sg (K.-K. Ma).

http://dx.doi.org/10.1016/j.jvcir.2016.03.016
1047-3203/

coding efficiency, Chen and Tsai [8] proposed a *joint temporal and spatial* MDC. Hsiao and Tsai [9] suggested a *Hybrid*-MDC framework by segmenting the input video in both spatial and frequency domains. To more effectively allocate redundancy, Bai et al. [10] developed a multiple description video codec by utilizing the correlation of the inter-/intra-descriptions at both the frame level and the *macroblock* (MB) level. Lin et al. [11] proposed a redundancy controllable H.264/AVC-based multiple description video coding algorithm. In addition, multiple description coding methods based on MV were proposed in [12,13]. Wen et al. [14] proposed an improved H.264-based drift-free multi-state MDC method by compressing the original video into multiple independent H.264/AVC streams with different coding parameters so as to control correlations between the descriptions.

In this paper, a more efficient MDC method, called the *adaptive data reuse* MDC (ADR-MDC), is proposed for robust video transmission. The proposed ADR-MDC method delivers a large improvement of our previous works as presented in [5,15]. In the proposed method, the input video sequence is first horizontally and vertically down-sampled by a factor of two on each frame to generate four sub-sequences, followed by grouping them into two descriptions via the quincunx manner. In each description, one sub-sequence is directly encoded and decoded by applying the H.264 codec, while the other is evaluated at each MB using the following adaptive data reuse criterion developed based on the spatial and temporal correlation between two sub-sequences: (1) If the corresponding MB in the directly-encoded sub-sequence (called as *co-located MB*) is considered as locating in a homogeneous, motionless or slow motion region, the encoding process of the current MB will be skipped, and the reconstructed co-located MB will be directly reused (namely, copied) in the decoding side; (2) if the co-located MB is considered as locating in the fast motion region (i.e., exploiting the small block-sized mode as the optimal mode), the same optimal mode and MVs of the co-located MB will be directly reused to perform the inter prediction and only the corresponding residual will be encoded, transmitted and decoded; (3) if the co-located MB is considered as locating in the highly-textured region (i.e., exploiting the Intra prediction mode as its optimal mode), the same optimal mode of the co-located MB will be directly reused to perform the intra prediction and only the corresponding prediction error will be encoded, transmitted and decoded. Therefore, the proposed ADR-MDC is able to achieve a more robust video transmission with higher coding performance and lower computational complexity, compared with the existing state-of-the-art MDC method.

The remaining sections of this paper are organized as follows. Section 2 describes the proposed ADR-MDC algorithm for robust video transmission in detail. Section 3 presents extensive experimental results to demonstrate the superior performance of the proposed ADR-MDC algorithm. Section 4 concludes this paper.

2. Proposed adaptive data reuse MDC (ADR-MDC)

2.1. Motivation

H.264/AVC adopts multiple sophisticated coding techniques, such as variable block sizes, multiple intra prediction modes, *rate-distortion optimization* (RDO), exhaustive mode decision, and so on, to achieve much higher coding efficiency. Due to its high coding performance, H.264 has been the most widely-used video coding standard in various video services and applications, such as IPTV, HDTV, 3DTV, to name a few. Therefore, we exploit the H.264/AVC as the platform for realizing the proposed ADR-MDC in this work.

Different from its predecessors, the H.264/AVC provides various prediction modes to fully adapt to various kinds of video contents for achieving much higher coding efficiency [16]. Variable block size *motion estimation* (ME) is used to exploit the temporal correlation inherited in video. The above-mentioned variable block sizes are 16 × 16, 16 × 8, 8 × 16, 8 × 8, 8 × 4, 4 × 8, and 4 × 4, and the last four block sizes are jointly denoted as P8 × 8 in the H.264/AVC, as shown in Fig. 1 [17]. Meanwhile, sophisticated intra prediction [18], including intra 4 × 4, intra 8 × 8, and intra 16 × 16, is used to remove the spatial redundancy. Accordingly, for inter-frame coding, there are 11 candidate modes: SKIP, 16 × 16, 16 × 8, 8 × 16, 8 × 8, 8 × 4, 4 × 8, 4 × 4, intra 4 × 4, intra 8 × 8, and intra 16 × 16. For the intra-frame coding, only intra 4 × 4, intra 8 × 8, and intra 16 × 16 are applicable. Among them, SKIP mode is more fit for the motionless region, a large block-sized mode (i.e., 16 × 16, 16 × 8, 8 × 16) is more suitable for a homogeneous region under slow motion, a small block-sized mode (i.e., P8 × 8) is more suitable for a region containing a fast moving object, and a intra prediction mode is more beneficial for a highly-textured region.

Generally speaking, the motion activities of the MB are intuitively related to its optimal mode [17]. In other words, all the above-mentioned prediction modes can be divided into three classes according to the types of motion activities, as shown in Table 1. To be more specific, Class 1 indicates the corresponding MBs locate in a motionless or slow motion region, Class 2 indicates the corresponding MBs locate in a fast motion region, and Class 3 indicates the corresponding MBs locate in a highly-textured or inhomogeneous region. Note that the proposed ADR-MDC method exploits a polyphase spatial subsampled method [4] to individually horizontally and vertically down-sample each frame of the input video sequence by a factor of two to generate four sub-sequences, as shown in Fig. 2. Hence, an MB in one sub-sequence indeed has little difference from the corresponding MB (i.e., at the same spatial position, called as co-located MB) of the other three sub-sequences in this case. Consequently, it is expected that there are strong correlations presented among the four sub-sequences. This means that the coding information (i.e., the optimal mode, MV, etc.) of the co-located MBs in four sub-sequences also has high correlation. To verify this observation, extensive experiments are conducted to study the correlation of coding information among four sub-sequences based on H.264/AVC and a set of test sequences. In our experiments, each test sequence is firstly sub-sampled into four sub-sequences as shown in Fig. 2 and each sub-sequence is encoded by H.264/AVC. Table 2 demonstrates the percentage of the co-located MBs in four sub-sequences that have the same optimal mode belong to Class 1, Class 2, and Class 3 under various QP values, respectively. From the results, it can be found that the co-located MBs indeed have high possibility to obtain the same coding information (e.g., the same optimal mode) – more than 80% for Class 1, 73% for Class 2, and 91% for Class 3. Based on the above analysis, we can directly encode one sub-sequence by reusing the coding information of other encoded

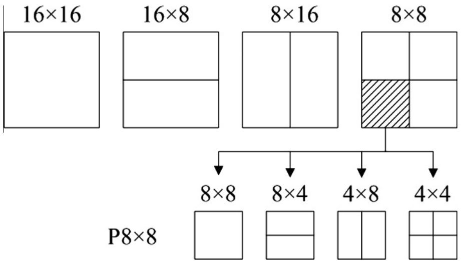

Fig. 1. An illustration of variable block sizes for ME in H.264/AVC.