J. Vis. Commun. Image R. 34 (2016) 78–88

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A fast mode decision algorithm applied in medium-grain quality scalable video coding ☆

Dayong Wang [a,e,*], Yu Sun [b], Yuanyuan Huang [c], Mingze Bai [a], Hongjian Li [d]

[a] School of Bio-information, Chongqing University of Posts and Telecommunications, Chongqing, China
[b] Department of Computer Science, The University of Central Arkansas, Conway, USA
[c] Department of Network Engineering, Chengdu University of Information Technology, Chengdu, China
[d] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China
[e] Graduate School at Shenzhen, Tsinghua University, Shenzhen, China

ARTICLE INFO

Article history:
Received 21 November 2014
Accepted 6 October 2015
Available online 19 October 2015

Keywords:
Temporal correlation
Spatial correlation
Inter-layer correlation
Correlation degree
Mode prediction
RD cost
Early termination
MGS coding

ABSTRACT

To increase the flexibility of bit stream adaptation in the H.264/Scalable Video Coding, Medium-Grain Scalable Video Coding (MGS), a variation of Coarse-Grain Scalable Video Coding (CGS), is included. The encoding structure of MGS is significantly different from that of CGS, and the encoding procedure of MGS is very complex. It is important to reduce encoding complexity without sacrificing MGS coding performance. In this paper, a fast mode decision algorithm is proposed for the enhancement layer of MGS based on its specific coding structure. First, the candidate modes and the coding sequence of the current macro-block (MB) are predicted based on mode correlations and correlation degrees. Next, early termination strategies, including Direct Mode, inter-layer, layer and spatial, and spatial-only, are proposed based on the coding structure of MGS and correlations. Finally, MBs are encoded in the order predicted with the four proposed early termination strategies to improve the coding speed. Experimental results show that the algorithm can achieve an average time saving of up to 85.44% with strong robustness and negligible loss in coding efficiency.

1. Introduction

As an extension of H.264/Advanced Video Coding (AVC), H.264/Scalable Video Coding (SVC) [1], can produce a single bit stream to adapt to various device capabilities, network conditions, and client applications. In H.264/SVC, there are temporal, spatial, and quality scalabilities that produce bit streams of scalable frame rates, image resolutions, and video qualities. To achieve spatial and quality scalability, SVC provides an H.264-compatible base layer (BL) and multiple enhancement layers (ELs). To promote the rate distortion (RD) performance of ELs, inter-layer prediction is added. Basically, all such features make the encoding procedure very complex. Therefore, developing a fast mode decision algorithm to reduce encoding complexity is very important to promote the wide application of SVC, particularly for mobile and real-time applications. In quality scalability, there are Medium-Grain Scalable Video Coding (MGS) and Coarse-Grain Scalable Video Coding (CGS). In CGS, scalability is achieved based on multiple CGS layers. On the one hand, too many layers will seriously degrade coding efficiency; on the other hand, too few layers will lead to a coarse granularity in scalability, which is not suitable in practical applications. To solve this problem, MGS was developed based on the so-called key frame concept and the flexible switching between different MGS layers [2]. With the MGS scheme, bit streams can be adjusted according to network conditions, resulting in strong adaptability and practical significance. It is therefore important to reduce the coding complexity of MGS.

There are a number of fast algorithms designed for accelerating CGS coding. Li et al. [3,4] and Lin et al. [5] predicted candidate modes and excluded low possible modes in EL based on the mode of the co-located macro-block (MB) in BL to improve the coding speed. Since the algorithm used inter-layer correlation to remove low possible modes, the coding speed is improved. Peng and Hang [6] predicted candidate modes in EL based on the mode of the co-located MB in BL. In addition, a BL reference frame index was selectively reused, and a BL motion vector was used as the initial search point for EL motion estimation (ME). Since the algorithm used more relative information in BL in prediction, the coding speed is further improved. However, the algorithm only used

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author at: School of Bio-information, Chongqing University of Posts and Telecommunications, Chongqing, China.
E-mail address: 18025441839@126.com (D. Wang).

http://dx.doi.org/10.1016/j.jvcir.2015.10.002
1047-3203/

inter-layer correlation to predict, the improvement of coding speed is limited. Shen et al. [7] proposed an efficient SKIP mode detection approach for coarse grain quality SVC. It made use of the coding information of spatial neighboring MBs in EL and the co-located MB in BL to predict SKIP mode MB. The algorithm used inter-layer and spatial correlations to predict, but the proportion of SKIP mode was limited after all, thus the coding speed was improved limitedly. Kim et al. [8] predicted candidate modes of the current MB in EL using the modes of the co-located MB and its neighboring MBs in BL. Since the algorithm fully used inter-layer and spatial correlations in MB mode to exclude low possible modes, the coding speed was improved obviously. Shen and Zhang [9] identified and utilized inter-layer and spatial correlations, the motion and mode characteristics of the current MB to adjust each step of ME in EL, including mode decision, search-range selection, and prediction direction selection. Since the algorithm used more information, such as inter-layer correlation, spatial correlation and the feature of current MB, the coding speed was further improved. These algorithms above basically used inter-layer and spatial correlations to exclude the low possible modes to improve the coding speed. However, how to accurately predict low possible modes is a hard work lead the coding speed improve not very significantly or coding efficiency. Generally speaking, early terminations are also efficient ways to improve the coding speed. Park et al. [10] firstly statistically derived the expectation of the rate distortion cost (RD cost) increase caused by skipping each mode in the mode decision; and then encoded the modes according to the expectation when the RD cost was smaller than the set threshold the encoding procedure was terminated. The algorithm used inter-layer correlation to predict and used the RD cost to early terminate, the coding speed was improved. However, the algorithm is developed only based on inter-layer correlation, so the coding speed was not improved very significantly. Yeh et al. [11] presented a fast mode decision algorithm that sped up the SVC encoding process through probabilistic analysis. The mode of EL was first predicted by statistical analysis. Afterwards, the Bayesian theorem was utilized to detect whether the prediction mode of current MB was the best or not. The mode was further predicted and refined by the Markov process. This algorithm used spatial correlation to predict and early terminate by the Bayesian theorem and the Markov process, the coding speed was improved. However, the algorithm is developed only based on spatial correlation, so the coding speed was also not improved very significantly. Jung et al. [12] proposed a fast mode decision algorithm for SVC that used all-zero blocks (AZB) detection. Based on empirical analysis of the inter-layer and spatial correlation of the AZB, MBs in EL was predicted to be AZB. Then, only predicted MBs were examined and terminated by the AZB detection algorithm. Although the algorithm used inter-layer and spatial correlations to predict and used AZB detection algorithm to early terminate, the proportion of AZBs was limited after all, the coding speed was also improved limitedly. Shen et al. [13] proposed an adaptive early termination of the fast mode decision algorithm in SVC. It made use of the RD cost of spatial neighboring MBs and the corresponding MB in BL to early terminate the mode decision procedure. The algorithm used inter-layer and spatial correlations in RD cost to early terminate, but only these correlations in RD cost are used, so correlations was not used very fully and the coding speed was not improved very significantly. Zhao et al. [14] developed a constrained model with optimal stopping based on inter-layer and spatial correlations, and then employed the solutions to this model to initialize the candidate mode list and early terminate the coding process. Since the algorithm fully used inter-layer, spatial correlations and optimal stopping, the coding speed was improved significantly. Lu and Martin [15] exploited inter-layer and neighboring correlations and examined the level of picture details and motion activity to predict candidate modes and early terminate the encoding process. Since the algorithm fully used more information including inter-layer and spatial correlations, level of picture details and motion activity to predict and early terminate, the coding speed was also improved significantly.

Although these algorithms are designed for CGS and also are applied for MGS, the coding structures of CGS and MGS are shown in Fig. 1. In CGS structure, all frames use the reference in the same layer for motion-compensated prediction. However, In MGS structure, all non-key frames marked by blue boxes typically use the reference with the highest available quality for motion-compensated prediction [16]. This means that non-key frames of MGS in EL and the co-located frames in BL are always predicted by the same reference frames (namely, the highest available quality frames), so the MBs in EL and the co-located MBs in BL are always predicted by the same reference frames. Obviously, the coding structures of CGS and MGS are different. It would be better to develop a fast mode decision algorithm based on the specific coding structure. However, to the best of our knowledge, no special fast algorithm for MGS has yet been proposed.

To the best of our knowledge, it is one of the pioneering researches that focuses on MGS fast encoding. In this paper, aiming at improving the coding efficiency, we proposed a fast mode decision algorithm for medium-grain quality scalable video coding. According to the mode correlations and correlation degrees, we proposed efficient methods to predict the candidate modes and the possibilities of the current macro-block (MB). We also developed early termination strategies based on the coding structure of MGS and correlations, and then we obtain early termination conditions by combining mode possibilities with complexities on that basis. To improve the coding speed, our algorithm encodes MBs in the order predicted by using the four proposed early termination strategies. The novelties of this algorithm includes: (1) employing both mode correlations and correlation degrees jointly to predict candidate modes and possibilities; (2) developing early termination strategies based on the specific coding structure of MGS; and (3) combining mode possibilities and complexities to obtain early termination conditions. Experimental results have demonstrated that the algorithm can improve coding speed significantly with negligible coding loss and high stability.

The rest of this paper is organized as follows: Section 2 introduces the mode decision procedure; Section 3 introduces the early termination strategies; Section 4 introduces decision of Inter8 $\times$ 8 and Intra Mode; and Section 5 presents the overall algorithm. The experiments are illustrated in Section 6, and conclusions follow in Section 7.

2. Proposed mode decision

In the proposed fast algorithm for MGS, the candidate modes and probabilities to be selected by the current MB in EL are predicted by exploiting temporal, spatial and inter-layer correlations. Generally speaking, other algorithms predict candidate modes only consider mode correlations without considering correlation degrees. Correlation degrees are closely related to mode prediction, so it is better to use correlation degrees and mode correlations jointly to predict candidate modes.

Fig. 2 shows the MBs for the prediction of the current MB mode. C is the current MB in EL; L is its left MB; U is its upper MB; UL is its upper-left MB; and UR is its upper-right MB. Accordingly, FC, FL, FU, FUL, FUR are the co-located MB of C, L, U, UL and UR in the previuos frame (i.e., the frame just encoded) of the current frame, respectively; BC, BL, BU, BUL, BUR are the co-located MB of C, L, U, UL and UR in BL, respectively; and FBC is the co-located MB of FC in BL.