J. Vis. Commun. Image R. 25 (2014) 1631–1639

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A fast mode decision algorithm applied to Coarse-Grain quality Scalable Video Coding

Dayong Wang [a], Chun Yuan [a,*], Yu Sun [b], Jian Zhang [c], Xin Jin [a]

[a] *Graduate School at Shenzhen, Tsinghua University, Shenzhen, China*
[b] *Department of Computer Science, The University of Central Arkansas, Conway, USA*
[c] *Schools of Software, University of Technology, Sydney, Australia*

ARTICLE INFO

Article history:
Received 22 January 2014
Accepted 24 July 2014
Available online 24 August 2014

Keywords:
Inter-layer correlation
Spatial correlation
Early termination
CGS coding
Scalable Video Coding
Quality SVC
Fast mode decision
Fast algorithm

ABSTRACT

A fast mode decision algorithm is proposed for a Coarse-Grain Scalable (CGS) video encoder based on the encoding characteristics of quality Scalable Video Coding (SVC). First, candidate modes and coding orders are predicted, based on inter-layer and spatial correlations. Three early termination methods are then proposed based on CGS encoding structure. Finally, all candidate modes are checked sequentially, according to their predicted order with three early termination conditions, to improve the coding speed. Experimental results have demonstrated that the proposed algorithm could reduce the encoding time by an average of 84.39%, with negligible coding efficiency losses.

1. Introduction

SVC is an extension of H.264/Advanced Video Coding (AVC) [1], which can produce a single bit stream that can adapt to various device capabilities, network conditions, and client applications. In SVC, three types of scalability are supported, including temporal scalability, spatial scalability, and quality scalability. To achieve spatial and quality scalability, SVC provides an H.264-compatible Base Layer (BL) and multiple Enhancement Layers (EL). To maintain coding efficiency comparable to that of single-layer video coding, inter-layer prediction was added to the process. The coding structure introduces high complexity to the encoder. Therefore, it is desirable to develop algorithms to reduce the coding complexity of the SVC, and to maintain the coding efficiency for many different applications, particularly for wireless and real-time applications. To reduce the coding complexity, a number of fast algorithms have been developed for quality Scalable Video Coding. Li et al. [2,3] and Lin et al. [4] employed mode relationships between the BL and the EL to reduce the number of candidate modes and to save encoding time. Kim et al. [5] predicted modes of each MacroBlock (MB) in the EL by using the modes of the co-located MB and its neighboring MBs in the BL. Peng and Hang et al. [6] optimized the mode decision for temporal and Coarse-Grain Scalable (CGS) coding, through a lookup table with the BL coding mode. In addition, the BL reference frame index was selectively reused and the BL motion vector was used as the initial search point for the EL Motion Estimation (ME). Shen et al. [7] proposed a SKIP mode detection approach for CGS. It made use of the coding information of spatial neighboring MBs, and the co-located MB in the BL, to predict SKIP mode MBs. Shen and Zhang [8] identified and utilized motion and mode characteristics of the current MBs, based on the inter-layer and spatial correlations, to adjust each step of the ME in the EL, including mode decision, search-range selection, and prediction direction selection. Park et al. [9] proposed a fast mode decision algorithm. It initially statistically derived the Rate-Distortion (RD) cost expectations due to mode skipping made by the mode decision, and then the encoder performed mode decisions by using a small number of modes that were determined based on the expected increase of RD costs. Jung et al. [10] proposed a fast mode decision algorithm for SVC which used All-Zero Block (AZB) detection. Based on empirical analysis of inter-layer correlations of an AZB, a prediction was made of whether the MB in the EL would be an AZB. Then, only predicted MBs were examined and terminated by the AZB detection algorithm. Yeh et al. [11] presented a fast mode decision algorithm that speeded up SVC encoding process through probabilistic analysis. A mode in EL was firstly predicted by statistical analysis. Afterwards, Bayesian theorem was

---

* Corresponding author.

http://dx.doi.org/10.1016/j.jvcir.2014.07.006
1047-3203/

utilized to detect whether the prediction mode of current MB was the best or not. The mode was further predicted and refined by Markov process. Shen et al. [12] proposed an adaptive early termination of fast mode decision algorithm in SVC. It made use of the coding information of spatial neighboring MBs in EL and the corresponding MBs in BL to early terminate mode decision procedure. Zhao et al. [13] developed a constrained model with optimal stopping, and the solutions to this model were then employed to initialize the candidate mode list and predict early termination. Lu and Martin [14] exploited the inter-layer and neighboring correlations, and examined the level of the picture details and motion activity, to make faster EL decisions and thus reduce the coding time.

All these research helped to improve coding speeds to some extent, but they did not take into consideration the encoding structure of CGS. In CGS, MBs in the EL and the co-located MBs in the BL are the same in content and coding parameters, except the Quantization Parameters (QP). To address this need, a fast mode decision algorithm for H.264 CGS coding has been proposed in this paper. The candidate modes and coding orders are predicted, and three early terminations are proposed in the EL, based on the CGS encoding structure. The experimental results have demonstrated that the proposed algorithm can reduce computational complexity significantly with negligible PSNR losses and bit-rate increases.

The remainder of this paper is organized as follows. Section 2 introduces the mode decision procedure, while Section 3 elaborates on the early terminations and Section 4 presents the overall algorithm. The experiments are illustrated in Section 5 and followed by the conclusions in Section 6.

2. The proposed mode decision procedure

Based on the structure's characteristics in CGS coding, a fast mode decision algorithm in the EL was proposed. In quality SVC, frames in the EL and BL have the same resolution, which means the MBs in the EL and co-located MBs in the BL have a one-to-one correspondence. Fig. 1 shows the MBs for the prediction of the current MB mode. C is the current MB in the EL, L is the left MB, U is the upper MB, UL is the upper-left MB, and UR is the upper-right MB. Accordingly, BC, BL, BU, BUL and BUR are the co-located MBs of C, L, U, UL and UR in the BL, respectively.

2.1. Spatial correlation

Since strong spatial correlations exist in natural video content, neighboring MBs are highly similar in motion and textural features, which results in a high correlation of their coding modes. The coding mode of the current MB can be predicted from its neighboring MBs, but it is not optimal to predict the current MB mode directly from the neighboring MB modes. The degree of correlation between the current MB and the neighboring MBs should be obtained first, and then the candidated modes should be obtained, and checked sequentially from those having the highest probability to those having the least probability. How the correlation degrees of the MBs are obtained in the EL is thus crucial. Since MBs in the BL and the co-located MBs in the EL are the same in content, the spatial correlations of the MBs in the BL can be utilized by the co-located MBs in the EL. Thus, the correlation degrees of the MBs in the BL should be obtained first, as described below.

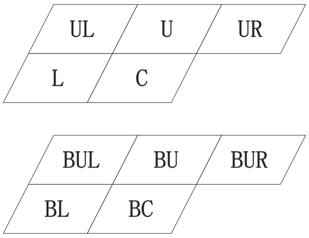

Fig. 1. MBs for the prediction of the current MB's mode.

Obviously, the more similar the modes of two neighboring MBs in the BL are, the more likely the modes of the co-located MBs in the EL will use the same modes. Therefore, the mode correlation between two neighboring MBs in the BL can be measured by a conditional probability that two neighboring MBs would select the same mode in the EL, given that the modes of their co-located MBs in the BL are m and n, respectively. It is given by the formula below,

$$p(\mathrm{MC} = \mathrm{MNC} | \mathrm{MBC} = m,\ \mathrm{MNBC} = n), \tag{1}$$

where MC and MNC denote the modes of the MB "C", and its neighboring MB in the EL, respectively. MBC and MNBC refer to the modes of their co-located MBs in the BL. In order to obtain the order of mode correlations, Akyio with CIF is selected to tested.

Based on statistical data from extensive experiments, the orders of the mode correlations were given in Table 1, where 16×16, 16×8, 8×16 and 8×8 referred to the mode of Inter16×16, Inter16×8, Inter8×16 and Inter8×8, respectively. The mode sequence from left to right shows the correlation order from high to low for all Inter modes under an MBC. When MBC is Direct mode, the most similar mode is Direct mode, the second most similar mode is Inter16×16, followed by Inter16×8 and Inter8×16, and the least similar mode is Inter8×8. Extensive experiments have showed that other sequences also follow the same rule. Namely, mode correlations have nothing to do with sequence.

From Table 1, it is obvious that the more similar the modes were, the stronger the correlations would be. For computational convenience in the next steps, the mode correlation degree d is represented as the mode correlation order under an MBC. Obviously, a larger d means a stronger correlation.

The above prediction is based on modes. If neighboring MBs in the BL use the same modes, their co-located MBs in the EL tend to use the same modes, and vice versa. However, since QPs in the BL and EL are different even if two neighboring MBs in the BL use the same mode, their co-located MBs in the EL could also use different modes. If two neighboring MBs use the same modes in the BL, but their co-located MBs use different modes in the EL, it is impossible to determine the degree of correlation of the two MBs in the BL by mode correlation alone. Therefore, the modes and rate distortion values of neighboring MBs would need to be used jointly to determine the MBs correlation in the BL. Since an MB could have a maximum of four neighboring MBs, the mode correlation degree d should be multiplied by 4 to guarantee the importance of the influence of the mode correlation over the Rate Distortion (RD) correlation, thus the mode correlation weight denoted as w^m is defined as follows:

$$w^m = 4d \tag{2}$$

The RD weight w^r is defined as the order in the *RD* value of neighboring MBs, relative to the *RD* value of the MB "BC". Thus, $w^r = 0$ is assigned to the mode whose *RD* value is the most different from the MB "BC"; $w^r = 1$ is assigned to the second most different, and so forth. Since both mode correlation and RD correlation have strong relationship with spatial correlation, the spatial weights of neighboring MBs in the BL are derived as described below.

$$w^s = w^m + w^r \tag{3}$$