LOW COMPLEXITY ALGORITHM FOR INTER-LAYER PREDICTION OF H.264/SVC

TIAN SONG, KENTARO TAKEI AND TAKASHI SHIMAMOTO
Graduate School of Advanced Technology and Science
The University of Tokushima
Minami-Jyosanjima 2-1, Tokushima City 770-8506, Japan
{tiansong; kentaro-t; simamoto}@ee.tokushima-u.ac.jp

Received March 2011; revised August 2011

ABSTRACT. Inter-layer prediction of H.264/SVC can help to improve the coding efficiency with increased computation complexity. This paper presents a low complexity algorithm for inter-layer prediction. The proposed algorithm focuses on the reduction of the candidate modes by making use of the correlations of the encoding cost between the base layer and enhancement layers. For macroblocks with the collocated macroblock coded by INTRA or INTER type two algorithms are proposed, respectively. Both algorithms efficiently decrease the redundant candidate modes by estimating from the base layer coding information. The experiment results show that the proposed algorithm can significantly reduce redundant computation complexity with almost no coding efficiency loss.

Keywords: H.264, Scalable video coding, Hierarchical B-picture, Search range

1. Introduction. In recent years, encoding standard which can achieve the scalability has increasing requirement due to the diversification of the network environment and the applications. Enhancing to the successful H.264/AVC, a scalable extension is standardized as H.264/SVC [1] in 2007. A reference software is also developed by the joint video team (JVT) for SVC [2, 3]. The objective of H.264/SVC is to enable the generation of a unique bitstream that can adapt to various bit-rate, transmission channel and display capabilities. In H.264/SVC, three scalabilities: spatial scalability, temporal scalability and quality scalability are finally recommended in the final draft [4, 5, 6]. However, due to its high implementation complexity, the complexity reduction becomes an important research issue. Some previous works are proposed to reduce the complexity of temporal scalability by adaptively reducing the redundant encoding modes or efficiently constructing the GOP (Group of Pictures) [7, 8, 9].

The resolution diversity of current display devices motivates the improvement for spatial scalability. The spatial scalability is realized by introducing multiple display resolutions within a single bit-stream. Therefore, the information of the input sequences and the selected modes in the base layer can be used to estimate the optimal mode in the enhancement layers, which is called inter-layer coding [3]. In the inter-layer prediction, three new modes have been introduced using the motion vectors, residuals, and intra information from the base layer to select the best coding mode in the enhancement layers. Using these new modes, inter-layer predictions can not only achieve scalable features but also improve the coding efficiency.

However, the inter-layer modes have to perform multiple times rate-distortion optimization (RDO) process, by which very high computational complexity is induced. In particular, the residual prediction mode has to perform twice of the RDO process which