Signal Processing: *Image Communication* 29 (2014) 1063–1078

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Fast macroblock encoding algorithm based on rate-distortion activity for multiview video coding

Wei Zhu [a,1], Yayu Zheng [a,2], Peng Chen [a,3], Jie Feng [b,*]

[a] College of Information Engineering, Zhejiang University of Technology, No. 288 Liuhe Road, Hangzhou 310023, China
[b] School of Informatics and Electronics & Electronics, Zhejiang Sci-Tech University, No. 5 the 2nd Street, Xiasha District, Hangzhou 310018, China

ARTICLE INFO

Article history:
Received 14 January 2014
Received in revised form
29 May 2014
Accepted 1 June 2014
Available online 10 June 2014

Keywords:
Multiview video coding
Mode decision
Motion estimation
Disparity estimation

ABSTRACT

Multiview video coding (MVC) is the appendix H of H.264/AVC, and it requires a great amount of time to compress multiple viewpoints' video with complex prediction structures. To reduce the whole computational complexity of MVC, this paper proposes a fast macroblock (MB) encoding algorithm based on rate-distortion (RD) activity, and it includes the fast mode decision and the fast motion/disparity estimation. First, the RD activity type of the current MB is calculated by utilizing the Skip/Direct RD cost and the average RD costs of classified MB modes. Then, through utilizing the RD activity type and RD costs of the estimated modes, the selection of candidate modes, the early decision of Skip/Direct mode, and the reduction of Inter8 × 8 mode estimation are all presented in the fast mode decision. By using the RD activity type and the correlations of vectors, the selection of search center and the prediction of search range are introduced in the fast motion/disparity estimation. In addition, the proposed algorithm can be applied to temporal and inter-view views as well as anchor and non-anchor frames. An experiment with a wide range of video scenes, camera setups and quantization parameters was implemented, and the results confirmed that the proposed algorithm can reduce the encoding time significantly while maintaining a similar RD performance as the original MVC encoder. Compared to the state-of-the-art algorithms, the proposed algorithm also demonstrated better performances in the various test cases.

1. Introduction

Multiview video is captured by multiple cameras from different viewpoints, and it is useful for various multimedia applications, such as immersive teleconferencing, three-dimensional (3D) television and free viewpoint television. Multiview video coding (MVC) compresses multiview video for efficient storage and transmission, and it has been standardized as the appendix H of H.264/AVC. Like H.264/AVC, MVC adopts variable block-size inter and intra predictions for each view [1]. It not only employs the traditional motion estimation (ME) for temporal prediction but also adopts the disparity estimation (DE) for inter-view prediction [2]. Since the computational complexity of single view encoding in MVC is quite involved, the whole computational complexity is tremendously huge for encoding multiple viewpoints. Although the multiview video plus depth (MVD) [3] technique can be used to reduce the number of viewpoints, it also needs to encode multiple viewpoints using MVC for better 3D experience of occluded objects.

* Corresponding author. Tel.: +86 13588492036; fax: +86 57186843576.
E-mail addresses: weizhu@zjut.edu.cn (W. Zhu), yayuzheng@zjut.edu.cn (Y. Zheng), chenpeng@zjut.edu.cn (P. Chen), arlose@zstu.edu.cn (J. Feng).
[1] Tel.: +86 13685789073; fax: +86 57185290717.
[2] Tel.: +86 13067946162; fax: +86 57185290717.
[3] Tel.: +86 13819195905; fax: +86 57185290717.

http://dx.doi.org/10.1016/j.image.2014.06.001
0923-5965/

Therefore, the fast algorithms in MVC should be studied for promoting its practical applications.

Because mode decision, ME and DE (ME/DE) of macroblock (MB) encoding consume most of MVC's encoding time, many proposed fast algorithms have focused on these aspects [4–36]. Some fast algorithms employ rate-distortion (RD) properties of MB modes and correlations of coding information to reduce the complexity of mode decision. Their optimization points include early Skip/Direct mode decision [4–12,35], early termination model [16–18], candidate modes selection [8–16,19–21,32–34] and prediction direction selection [12–14,19,32,34]. Other fast algorithms adopt correlations of motion vectors, characteristics of disparity vectors, and the geometry relation between motion vectors and disparity vectors to reduce the complexity of ME/DE, and their optimization points include search range adjustment [21–28,32–34,36], multi-reference frames selection [24,29] and iterative search strategy [30,31]. To reduce the complexity of MVC in a MVD system, a depth map is used to select the inter-view prediction and the limited candidate modes for color video coding [19,20]. The video-depth correlation of coding information is employed to reduce the number of candidate modes and the search range of ME for depth map coding [11,21]. To reduce the whole complexity of MVC, several fast encoding algorithms address both the fast mode decision and fast ME/DE. In [32], the prediction direction is determined according to the coding information in the reference frame, and the search range of DE is reduced by using the location relationship between views, and candidate modes are selected based on the relationship between the temporal prediction and inter-view prediction. In [33], the search center and search range of ME are predicted by using inter-view correlation of motion vectors, and candidate modes are selected by utilizing the inter-view correlation of RD costs. In [34], Shen et al. proposed a view-adaptive ME/DE algorithm, where candidate modes, the search range of ME, and the inter-view prediction direction are jointly optimized based on mode complexity and motion homogeneity. In [35], Zhang et al. proposed an overall MB encoding algorithm which combines the early termination model with the fast multi-reference frame selection to reduce the encoding time for all views. In [36], Khattak et al. presented a complete low complexity MVC encoding solution (CLCMVC), which reduces the encoding time through the integration of early Skip/Direct mode decision, candidate modes selection, prediction direction selection, reference frame selection and search range reduction.

Generally speaking, above mentioned algorithms can reduce the complexity effectively, but the whole complexity of MVC is still huge. For sequences with fast motions and large disparities, either the computational performance or the RD performance is not very satisfactory. In addition, although some algorithms have considered the effect of quantization parameter (QP) variation, the stability of computational performance and RD performance under large range QPs can be further improved. In this paper, we aim to reduce the whole complexity of MVC significantly while maintaining a similar RD performance as the original MVC encoder under various test cases. For each MB encoding, both a fast mode decision and a fast ME/DE are addressed. The estimation of Skip/Direct mode is performed first to predict the RD activity of the current MB. Then the following fast mode decision and fast ME/DE are optimized by using the RD activity type, and the inter-view and spatial correlations of coding information are adopted to promote the performance.

The rest of the paper is organized as follows: Section 2 introduces the motivation and analysis, and Section 3 describes the proposed algorithm. The experimental results and analysis are given in Section 4, and conclusions are summarized in Section 5.

2. Motivation and analysis

In the MVC reference software JMVC, the candidate modes of each MB consist of inter and intra modes, and they are estimated to select the optimal MB mode. Inter modes include Skip/Direct, Inter16 × 16, Inter16 × 8, Inter8 × 16 and Inter8 × 8, where Inter8 × 8 further includes 8 × 8, 8 × 4, 4 × 8, and 4 × 4 sub-modes. Intra modes include Intra16 × 16, Intra8 × 8 and Intra4 × 4. Moreover, ME/DE are performed for inter modes excepting Skip/Direct, and the default search range of ME/DE in JMVC is up to 96, which is much larger than that in H.264/AVC reference software JM. The impact of numerous candidate modes and the large search range of ME/DE are two of the major reasons for its huge computational complexity, and they need to be optimized in order to reduce the whole complexity. Characteristics of MB modes and search range were analyzed in this section by employing the MVC reference software JMVC with version 8.0. The basic test conditions in this section are as follows: for each sequence, view 0 and view 2 are selected as reference views, and view 1 is used for statistical purposes. For each view, the GOP length was set to 12 and five GOPs were coded. For each frame, the default coding conditions were used: maximum two reference frames are available for the forward reference list and backward reference list respectively, one inter-view reference frame is allowed for each reference list, and the search method “TZ search” is enabled with search range 96.

Table 1 gives the proportions of MB modes under two typical basis QPs for different sequences. Basis QP24 was chosen for the statistics under a high bitrate and basis QP36 was chosen for the statistics under a low bitrate. Three 640 × 480 sequences (“Exit”, “Ballroom” and “Race1”), two 1024 × 768 sequences (“Breakdancers” and “Ballet”), and one 1280 × 960 sequence (“Dog”) were selected for statistics. It can be seen that Skip/Direct occupies the majority proportion of MB modes under both the low and high bitrates for different sequences, but its estimation consumes negligible complexity due to it being without ME/DE. Inter16 × 16 occupies the second proportion, and its average proportions are 18.5% and 12.3% under basis QP24 and QP36 respectively. Inter16 × 8/Inter8 × 16 occupies a considerable proportion, especially for sequences with large motion and disparity, such as “Ballroom”, “Race1” and “Breakdancers”. Inter8 × 8 occupies a small proportion, while it consumes more encoding time than the other modes because it includes four sub-modes. Intra modes have small proportions, and their complexities are smaller than that of the inter modes. Based on the above statistics and analysis, it can be shown that the proportions and the complexities are widely different among