J. Vis. Commun. Image R. 23 (2012) 522–534

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

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

Iterative search strategy with selective bi-directional prediction for low complexity multiview video coding

Zhi-Pin Deng [a,b], Yui-Lam Chan [a,*], Ke-Bin Jia [b], Chang-Hong Fu [a], Wan-Chi Siu [a]

[a] Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[b] Department of Electronic Information and Control Engineering, Beijing University of Technology, Beijing, China

ARTICLE INFO

Article history:
Received 14 July 2011
Accepted 16 January 2012
Available online 3 February 2012

Keywords:
3DTV
Multiview video coding
Disparity estimation
Motion estimation
Bi-directional prediction
Hierarchical B picture
Adaptive search range
JMVC

ABSTRACT

The multiview video coding (MVC) extension of H.264/AVC is the emerging standard for compression of impressive 3D and free-viewpoint video. The coding structure in MVC adopts motion and disparity estimation to exploit temporal and inter-view dependencies in MVC. It results in a considerable increase in encoding complexity. Most of the computational burden comes from uni-directional and bi-directional prediction. In this paper, an iterative search strategy is designed to speed up the uni-directional prediction in MVC. It can work with an adaptive search range adjustment through a confidence measure of a loop constraint to obtain both motion and disparity vectors jointly. Furthermore, a selective bi-directional prediction algorithm is proposed to enhance the coding performance by analyzing the statistical characteristics of bi-directional prediction in MVC. Experimental results demonstrate that, by using the proposed fast search, the temporal and inter-view redundancies of multiview video can be eliminated sufficiently with low complexity.

1. Introduction

With recent advances in stereo and three dimensional (3D) display technologies, 3D video has become an emerging medium that can offer a richer visual experience than traditional video [1]. Potential applications include free-viewpoint video (FVV), free-viewpoint television (FTV), 3D television (3DTV), IMAX theater, immersive teleconference, surveillance, etc. [2]. To support these applications, video systems require capturing a scene from different viewpoints which result in generating several video sequences from different cameras simultaneously.

Multiview video coding (MVC) [3] has been studied for a long time in the Joint Video Team (JVT) formed by ISO/IEC MPEG and ITU-T VCEG. An international standard on MVC was developed in July 2008 as the H.264/AVC Multiview High Profile [4]. It comes along with the reference software referred to as Joint Multiview Video Coding Model (JMVC) [5]. The JMVC supports multiple reference frame selection from either the same view or neighboring views. Take a three-view coding structure with a group-of-picture (GOP) length of 12 shown in Fig. 1 as an example. In this example, S_i denotes the ith view. There are two types of frames in this MVC prediction structure. Anchor frames (enclosed by dotted lines in Fig. 1) are placed at the beginning of a GOP while non-anchor frames lies in between two anchor frames. The JMVC uses the block-based motion estimation (ME) and disparity estimation (DE) to exploit both temporal and view correlation. Within the same view, the JMVC adopts hierarchical B picture coding (HBP) [6] as the basic temporal prediction structure such that the non-anchor frames of a GOP are classified into different temporal layers, denoted by TL_1, TL_2, TL_3, and TL_4. The non-anchor frames in the current temporal level are usually hierarchically predicted by referring the frames in the last temporal level. From Fig. 1, S_i is also coded by referring two neighboring views S_{i-1} and S_{i+1}, and is called as B-view. This prediction structure offers higher coding efficiency at the expense of dramatically increased computations. With this arrangement, the prediction process in each macroblock (MB) of a non-anchor frame includes forward and backward motion estimation, forward and backward disparity estimation [7] and bi-directional prediction [8]. The one with the minimum rate-distortion (R-D) cost is selected as the final prediction type of the current MB. The hybrid uni-directional prediction and bi-directional prediction schemes make the prediction process as the most computationally intensive part [9].

Many fast search algorithms have been proposed to reduce the computation time of the prediction process in MVC. Since MBs with slow/homogeneous motion prevail in a sequence and these MBs always prefer ME to DE, some early termination algorithms were

* Corresponding author. Fax: +852 23628439.
E-mail addresses: zhipindeng@gmail.com (Z.-P. Deng), enylchan@polyu.edu.hk (Y.-L. Chan), kebinj@bjut.edu.cn (K.-B. Jia), enchfu@hotmail.com (C.-H. Fu), enwcsiu@polyu.edu.hk (W.-C. Siu).

1047-3203/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2012.01.016

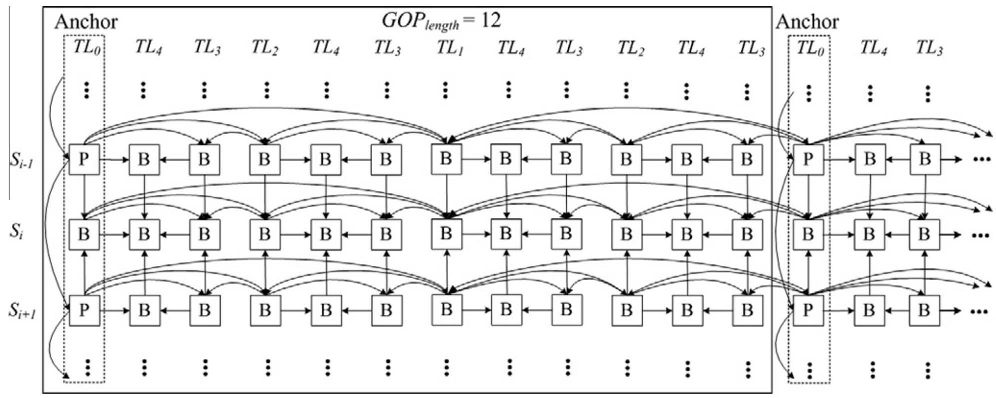

Fig. 1. MVC prediction structure using hierarchical B pictures.

proposed to selectively skip DE according to the rate-distortion (R-D) cost of ME or the characteristics of MBs/frames [10–13]. However, the determination of the stopping criterion is still a knotty problem. Based on the camera geometry and coding information of the corresponding MB in neighbor view, a reduced search range of ME/DE instead of exhaustive full search was suggested in [14]. However, additional information of multiview video such as the camera geometry is required and the performance might be well only for certain video sequences. Some predictor based fast ME/DE algorithms were then proposed [15–18]. With the aid of the already-known disparity vector, several predictors in ME can be obtained by tracking through the corresponding MB in the pre-coded inter-view reference frame. Then a fast ME algorithm with just a small search window is adopted to ensure the coding performance. Since the overall performance of ME is easily influenced by the accuracy of the previously estimated disparity vector, the full search is usually performed in DE prior to ME, in order to get an accurate disparity vector. Although these algorithms can speed up ME, the complexity of DE cannot be reduced simultaneously. Thus some algorithms for reducing the complexity of both ME and DE were proposed in [19–22]. These algorithms exploit the loop constraint among neighboring motion and disparity vectors to expedite the search process and remove some useless search region. Great complexity can be reduced with good coding performance. But in most of these methods, the ME and DE are relatively independent; the vector of motion/disparity field in the last step cannot be fully used to refine the vector of disparity/motion field in the next step. Moreover, all the above algorithms do not consider the bi-directional prediction in hierarchical B picture coding of MVC. In the JMVC codec, the bi-directional prediction is performed other than forward and backward predictions on all block sizes in order to search for a better result and eventually identify the final prediction type. Speeding up the bi-directional prediction must make a great contribution for the whole prediction part.

In this paper, we present a fast and efficient algorithm to reduce computational burden of both uni-directional prediction and bi-directional prediction. In our previous work [23], a fast prediction algorithm based on a loop constraint was proposed to expedite the uni-directional prediction by iteratively estimating motion and disparity vectors. A confidence measure of the loop constraint was then designed to adaptively adjust a search range for each iteration. Our previous algorithm only targeted to the uni-directional prediction. As a result, we could not achieve the optimal performance of the JMVC codec in which both uni-directional prediction and bi-directional prediction are included. In this paper, we further propose a new bi-directional prediction technique worked with our previous algorithm in order to further speed up the JMVC codec. The proposed selective bi-directional prediction is performed in hierarchical B picture coding of MVC based on the analysis of the prediction type selection process. Experimental results show that the proposed algorithm can reduce the computational complexity significantly compared with the conventional search and keep the coding quality.

The remainder of this paper is organized as follows. Section 2 gives the reliable uni-directional prediction through the confidence measure using the loop constraint. The selective bi-directional prediction technique is then presented in Section 3. Section 4 introduces the flowchart of the proposed algorithm. Simulation results and discussions are presented in Section 5. Finally, Section 6 gives a summary of the contribution of this paper.

2. Joint uni-directional prediction scheme

As shown in Fig. 2, a non-anchor frame in the B-view (S_i) has at most 4 reference frames from both temporal and view directions. Suppose the non-anchor frame at t in S_i is the current frame, denoted by $f_{i,t}$. In the JMVC codec, the forward reference frames $f_{i,t-T}$ and $f_{i-1,t}$ belong to list0. Similarly, the backward reference frames $f_{i,t+T}$ and $f_{i+1,t}$ belong to list1. In course of search, firstly, the motion and disparity vectors of uni-directional prediction are estimated independently. Four predictive vectors of the current MB ($MB_{i,t}$), including the forward motion and disparity vectors ($MV_{FW,i}$ and $DV_{FW,t}$, respectively), and the backward motion and disparity vectors ($MV_{BW,i}$ and $DV_{BW,t}$, respectively), are predicted independently under exhaustive motion or disparity estimation in each reference frame. Secondly, the bi-directional prediction is employed by iteratively searching from one past reference frame from list0 and one future reference frame from list1. At last, the best prediction type (forward, backward, or bi-directional) is determined by evaluating the one with minimum R-D cost. Although exhaustive independent

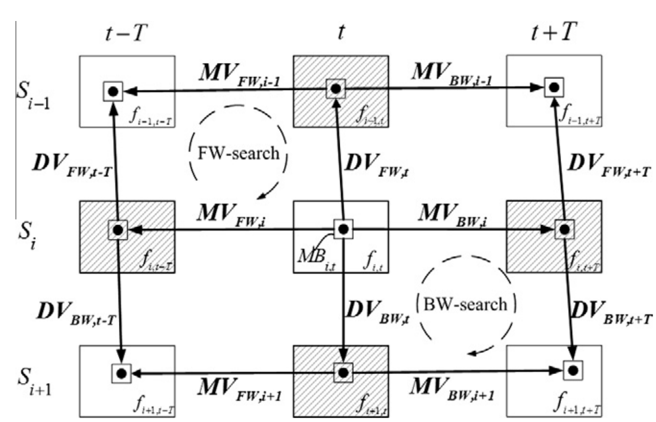

Fig. 2. Loop constraint.