J Real-Time Image Proc
DOI 10.1007/s11554-013-0328-3

ORIGINAL RESEARCH PAPER

Fast mode decision based on texture–depth correlation and motion prediction for multiview depth video coding

Zhaoqing Pan · Yun Zhang · Sam Kwong

Received: 20 September 2012 / Accepted: 8 February 2013

Abstract The motion estimation and disparity estimation are used to remove the temporal and inter-view redundancies in multiview plus depth video coding, however, the variable block-size ME and DE make the computational complexity increase dramatically. This drawback limits it to be applied in real-time applications. In this paper, based on the mode correlations between depth video and its corresponding texture video, motion prediction and coded block pattern, we propose a fast mode decision algorithm to reduce the computational complexity of multiview depth video coding. Experimental results show that the proposed algorithm can achieve 67.18 and 69.90 % encoding time saving for even and odd views, respectively, while maintaining a comparable rate-distortion performance. In addition, with the dramatic encoding time reduction, the proposed algorithm becomes more suitable for real-time applications.

Keywords Three-dimensional video · Fast mode decision · Multiview depth video coding · Video coding

Z. Pan · Y. Zhang · S. Kwong (✉)
Department of Computer Science,
City University of Hong Kong,
Kowloon, Hong Kong
e-mail: cssamk@cityu.edu.hk

Z. Pan
e-mail: zqpan3@student.cityu.edu.hk

Y. Zhang
e-mail: yunzhang@cityu.edu.hk

Y. Zhang
Shenzhen Institutes of Advanced Technology,
Chinese Academy of Sciences, Shenzhen, China

1 Introduction

As the demand for real-world visual perception increases, three-dimensional (3D) video is becoming more and more popular. Multivew video plus depth (MVD), which consists of multiview texture video and corresponding depth video, is an advanced 3D video representation format for the 3D applications, such as free-viewpoint television (FTV), three-dimensional television (3DTV) broadcasting, immersive teleconference and so on. The multiview texture video is captured simultaneously by multiple cameras from different viewpoints, and the depth maps provide the geometrical information for their corresponding texture video. At last, the 3D video is generated by image-based rendering techniques [1]. However, as the number of capturing cameras increases, the volume of raw texture and depth video data increases rapidly. To efficiently encode MVD, multiview video coding (MVC) is developed as an extension of H.264/AVC standard to exploit spatial, temporal and inter-view redundancies [2, 3]. Due to the variable block-size motion estimation (ME) [4] and disparity estimation (DE), the computational complexity of MVC is quite high.

To address the high computational complexity of variable block-size ME, a number of fast mode decision algorithms have been proposed for H.264. Based on the distribution of motion activity in each frame, a fast mode decision was proposed for H.264/AVC [5]. Hu et al. [6] proposed a fast inter mode decision for H.264/AVC, based on rate-distortion (R-D) cost characteristics. Zhao et al. [7] proposed an adaptive fast mode decision algorithm, which projects all candidate modes into a 2-D map, then the mode decision is performed according to a priority-based mode candidate list. Based on the optimal stopping theory and all-zero block detection, they also proposed a fast mode

Published online: 03 March 2013