摘要

立体电影的上马，譬如阿凡达的热映，引发了人们对立体电影的竞相追捧。立体电影的原理是利用人双眼的视角差和会聚功能，而产生了三维立体的效果。多视点视频技术，由于是立体电影的技术基础，而成为了近来研究的热点。因此，多视点视频压缩技术的研究意义重大，前景广阔，十分值得深入地研究。多视点视频生成是由多个摄像机按照一定地几何排布后，对着同一场景，采集而成。由于拍摄的场景大致相同，每个摄像机所拍摄的视频内容往往是表现了同一场景，或仅仅有些不同，比如：物体因为从不同角度拍摄，造成了几何形变扭曲而已。这些相似的大量信息，在时域、空域上存在了大量的冗余信息，于是给视频数据的压缩提供了可能。

目前解决的方式有很多种，包托利用当前的现有编码标准，比如：H.264/MPEG-4 AVC 去直接编码每个视点下的各路视频，也就是所谓的直接编码方式。除此之外，MPEG-2 提出了传统立体视编码的方法，也就是将两路对应于人眼距离的视频，利用时域和空域的依赖关系，压缩编码。目前最主要的方法是采用多视编码框架，由于层次 B 结构能获得更多的编码增益，因此 JVT 已经采用了层次 B 的框架作为标准。它的核心思想是分别运用运动补偿和视差补偿的方法，去掉时域、空域上的冗余信息。

虽然国际上的标准已经对立体视频研究了很多年，并且相继制定了很多标准，国内对此研究仍然存在很多空白。AVS 作为国内存有自主产权的核心编码标准，对于前沿的研究热点也责无旁贷地投入了研究工作。而本文的主要贡献也是在 AVS 标准上对适应立体编码标准，提出了相应的技术优化，具体包括如下两个方面：

第一，本文提出了支持 AVS 立体视频编的系统层技术方案，并被采纳进 AVS 标准。通过本文所提的提案，可以使得 AVS 标准在两个层次上实现对立体视频编码的支持。在 MPEG 系列国际标准中，虽然 MPEG-2 标准仅支持简单的视间预测编码， H.264 MVC Profile 能够支持基于视间预测的视频编码或者基于 SEI 的拼接编码，但对基于深度的立体视频编码没有任何考虑。相比之下，AVS 则在系统

和视频两个层次实现了对深度、视间预测等立体视频编码技术的支持。特别是若采用了双视点拼接编码框架，这种单一的多视视频流在视频层上与传统的视频流没有任何区别。在大量节约了码率的同时，也保证了观看质量，并且在实践阶段，此种方式已经提供给了电视观众立体的显示效果，让实时观看成为了可能。

第二，本文提出了一种自适应的运动跳过模式。此种运动跳过模式（motion skip）在多视框架上，MOTION SKIP 模式作为一种新的模式，被收录到研究模型当中。该模式参考附近视点序列中相关的宏块信息，包括：参考帧号，宏块编码模式，运动矢量，从而达到减少冗余信息的效果。然而，本文所提的方法，特别针对的是在 AVS 下采样辅助的双拼 MVC 编码框架。由于采用的是 AVS 编解码器，为了适应其最小粒度的编码单位，本算法采用了 8 像素精度的运动估计来获得参考图像到编码图像上的全局视差矢量。之后借助预定义窗口大小的细搜索过程，算法可以逐渐找到真实视差和 GDV 之间的误差。一旦得到最优偏移视差矢量，全局视差矢量会根据搜索结果不断完善优化自己以求达到最匹配真实视差矢量的效果。实验结果表明，通过改进 GDV 的获得方式，率失真性能得到了显着的提升。

综上所述，本文针对在 AVS 上实现立体编码问题，利用系统层上的改进，实现了双拼框架和高清实时编码的系统层设计。本文还提出了自适应的运动跳过模式，并通过实验展示了利用自适应地优化全局视差矢量，可以进一步提高编码效率。

关键词：多视编码，多视描述子，双拼多视编码框架，运动跳过模式，全局视差矢量。

Research on AVS 3D video coding

Lianlian Jiang (Computer Applied Technology)

Directed by Wen Gao

Three-dimensional movie, such as Avatar, has aroused people's desire for watching three-dimensional (3D) movie. The angular difference of each eye and convergence function can produce three-dimensional effect, which is the basic theory of 3D movie. Therefore, multi-view video technology, as the basis of three-dimensional film becomes a hot spot for researchers. The compression of multi-view video is the most significant technology for 3D program's system. Multi-view video is captured from different viewpoints of the same scene. The shooting scene is roughly the same, with merely some difference or distortion caused by different angles. Due to the similar scene, there are vast amount of redundant data to be encoded.

The way to solve it is to directly encode each sequence by current coding standard, such as H.264/MPEG-4 AVC video encoding standard. Later MPEG-2 added multi-view profile (MVP), as an amendment to the MPEG-2 standard, in order to encode stereo video. From then on, a number of coding technologies and standards, aiming at supporting 3D effect, have been proposed. JVT has adopted the framework of hierarchical B as multi-view video coding standard. The scheme uses both motion compensation and disparity compensation to further remove redundant data from temporal and spatial domain.

In this paper, we proposed a method in this thesis devoted to system layer of AVS stereo video coding standard, which was adopted into AVS video coding standard issued by Audio Video coding Standard (AVS) Workgroup of China. The method as improvement in system layer can enforce the three-dimensional effect at two levels by means of current broadcasting infrastructure. Although MPEG-2's multi-view profile is able to support stereo video coding based on the state-of-art forecasting infrastructure with the aid of SEI-based splicing syntax, but it doesn't take depth

video into consideration. Contrary to MPEG-2, AVS video coding standard can support both texture and depth video coding. Sampling-aide MVC scheme as one profile in AVS, without significant modification of current video coding standards, can merge two videos from neighboring viewpoints into one video. It not only saves a lot of bitrates, but also ensures viewing quality, which actually makes it possible to three-dimensional display with real-time coding and decoding.

Second, this thesis presents an adaptive motion skip mode. Inter-view motion skip mode has been proposed to improve the coding efficiency of multi-view video coding (MVC) by reusing the motion information of the referenced views. In this paper, an adaptive motion estimation algorithm for the motion skip mode is proposed for AVS 3D video coding. The proposed algorithm searches the best motion information, by means of adaptive global disparity estimation and a refine search along the horizontal direction, for the purpose of motion-compensated coding. Moreover, the method is applied to sampling-aided scheme for AVS 3D video coding, which encodes the reorganized sequence by merging two down-sampled videos. Rate-distortion optimization criterion is employed to find the best motion information. Experimental results demonstrate that the proposed algorithm can substantially improve the coding efficiency.

In summary, this thesis proposed two methods to improve three-dimensional video coding efficiency and experimentally demonstrate that the proposed approaches for multi-view video coding are feasible and effective. The improvements in the system layer of AVS assist to realize real-time HD multi-view video coding. The experiment result of adaptive motion skip mode shows the proposed method can adaptively optimize the global disparity vector and further improve coding efficiency.

Keywords: multi-view video coding, multi-view descriptor, Sampling-aide MVC scheme,adaptive motion skip mode, global disparity vector.