摘　要

随着采集和显示技术的发展，部署在工业和消费电子产业的多视点视频系统越来越多。多视点视频通过一组同步的摄像机阵列从不同角度采集同一场景，它是构成三维电视和自由视点视频等新型媒体应用的基础。由于多视点视频系统涉及到巨大的数据量，给存储和传输带来了巨大的挑战。为了存储和传输多视点视频，高效的多视点压缩是实现多视点视频系统应用的非常必要的工作。目前，它已经成为科研人员的研究热点。为了对多视点视频应用系统提供关键的技术保证，本论文重点解决多视点视频和深度序列压缩中的关键技术。针对面向三维视频和自由视点视频应用的多视点视频系统，本文从系统全局的角度着手，解决多视点视频系统中的压缩问题。根据系统中各个环节对压缩的需求，在保证系统需求的灵活性和适配性的同时，研究多视点视频和深度序列的高效编码算法。本文从提高压缩效率开始，分析随机访问与压缩的关系，研究压缩与视点绘制的适配关系，研究多视点视频和深度的联合码率控制。本文的主要研究内容如下：

第一，根据多视点编码中，关键帧位置的视点方向编码在整个多视点编码中编码性能贡献较大的特点，本文提出利用视点变形对齐的策略来提高视点方向编码的效率。根据多视点视频之间的几何关系，利用全局运动估计，估计出对齐模型的参数，然后把相邻视点的参考图像向当前编码图像对齐。利用多视点补绘技术对变形对齐的图像上产生的空洞进行补绘。补绘好的变形对齐图像被送入当前编码帧的参考帧缓冲区，作为编码的预测参考帧。为了进一步提高变形对齐的预测效率，本文使用多假设视差预测，形成联合优化的加权预测信号。实验证明，本文提出的这种方法可以有效的提高视点方向的编码效率。

第二，迅速的随机访问压缩码流中某个特定数据段是视频应用中交互功能的重要保证。在目前的基于 H.264/AVC 的多视点编码中，由于引入了视间预测，帧与帧之间的依赖严重，相应的也导致了随机访问码流的困难。为了解决压缩效率和随机访问之间的平衡问题，本文针对不同类型的随机访问，提出了相应的解决方法。针对时域的随机访问，本文提出了一种灵活调整预测结构的算法来提高时域方向的随机访问能力。根据随机访问的约束条件，可以利用该算法找到一个相对最优的预测结构，该结构能够保证压缩效率和随机访问之间保持最优的平衡。针对视点的随机访问，本文提出了调整预测结构和辅助表示编码的思想来提高视点的随机访问能力。在视点方向上本文提出了三种不同的编码结构：一是间隔交织预测的结构，二是采用切换帧的视点编码结构，三是基于辅助表示的编码结构。针对空域随机访问的定义，在基于 H.264/AVC 的多视点编码框架，本文提出了一种新颖的独立时空体编码方法，限制不同区域之间的相互预测，能够支持快速的空域数据随机访问。

第三，在多视点视频系统中，虚拟视点的绘制一般都是由压缩的视频和深度来完成

的。压缩对视点绘制会产生很大的影响。针对这一问题，本文对于压缩如何导致视点绘制失真这一问题进行了量化的分析。深度压缩损失会导致视点变形映射位置的偏移，而视频压缩的强度损失会映射到虚拟视点上。通过把视点绘制问题看作是数学上的线性插值问题，本文给出了压缩引起的视点绘制失真的理论上限，该上限主要取决于深度压缩引起的映射位置误差和视频压缩引起的强度损失。实验证明，提出的理论上限可以近似的反映虚拟视点绘制的质量。

第四，在多视点视频和深度压缩中，如何分配深度和视频之间的码率直接影响到虚拟视点的绘制质量。针对这一问题，本文通过理解视点绘制过程中压缩的影响，总结出视点绘制的失真模型。提出的视点绘制失真模型主要考虑了三种失真影响，即视频压缩引起的绘制失真，深度压缩量化引起的绘制失真，以及固有的几何失真。视频压缩引起的绘制失真可以通过计算直接得到，而深度压缩引起的绘制失真是通过一个基于视频功率谱的线性模型来刻画的。原始的几何失真是通过辅助的虚拟视点绘制进行估计的。根据提出的视点绘制失真模型，虚拟视点的绘制失真可以在没有原始视频作为参考的情况下得到有效的估计。在虚拟视点绘制失真模型的基础上，本文进一步提出了基于视点绘制模型的深度和视频的联合码率分配方法。该方法通过绘制失真模型评价不同码率分配组合时的虚拟视点失真来指导最优的码率分配。实验表明，在低码率端，和基于全搜索或者分层次搜索的码率分配方法相比，基于模型的方法花费相对较少的计算代价，达到几乎相同的码率分配效果。在高码率端，提出的基于模型的码率分配方法也可以节省巨大的计算而达到和分层次搜索的方法相同的码率分配效果。

第五，在三维多视点系统应用中，如何提供与信道带宽适配的信源编码码率是一个多视点编码必须解决的码率控制问题。传统的码率控制算法都是针对二维视频的，为了能够提供适配恒定信道带宽的三维视频服务，本文首先利用图像拼接的方式同时编码深度和视频，然后提出多视点深度和视频的联合码率控制算法。提出的码率控制算法包含三个层次的码率分配和控制。在视点这一层次，根据不同的视点编码预测类型，采用了不同比重的码率分配。在深度和视频这一层次，不均衡的分配深度和视频的目标码率，这样可以保证在一定的总码率约束下，视频可以得到较多的码率，从而保证较高的图像质量，后向兼容二维视频的显示应用。在帧一级码率控制中，针对分层次的 B 帧预测，使用了分层次的码率分配策略来为每一帧分配码率。在码率分配中除了考虑带宽的限制以外，本文还考虑了多视点视频输出时的缓冲区限制问题。提出了针对输出多视点视频时的兼顾多视点 HRD 的码率控制策略，可以避免多个视点同时输出时的缓冲区上溢或者下溢。实验结果显示，本文提出的码率控制算法可以满足三维视频系统应用，有效的控制码率和编码输出缓冲区的状态。

关键词：多视点视频，多视点深度，视频编码，随机访问，深度和视频联合码率分配，深度和视频联合码率控制

Research on Multiview Video and Depth Coding

Yanwei Liu (Computer Applications)
Directed by Qingming Huang

With the rapid developments of imaging and display technologies, more and more multiview video systems are deployed in the industry and consumer electronics field. Multiview video simultaneously records the same scene by a set of cameras from different viewpoint and it is a basic component of new media application such as 3DTV and free viewpoint video. The huge amount of multiple videos is a great challenge for storage or transmission. Multiview video coding is the essential work to reduce the data amount for realizing the multiview video system. Recently, it is becoming a hot subject of research. To provide the key technologies and feasible reference solution for multiview video system, this thesis focuses on resolving the key techniques for multiview video and depth coding. According to the specific application of 3D video and free viewpoint video, we understand multiview video and depth compression from the systematic viewpoint. According to the requirement of other components to the compression in the system, our work aims at seeking the compression's flexibility and adaptivity to the system besides improving the compression efficiency. This thesis starts from improving compression efficiency and then analysises the relation between compression and random access. After that the relation between compression and virtual view synthesis is discussed. Specifically, under the constraint of channel rate, how to jointly allocate the video rate and depth rate to guarantee the optimal view synthesis quality is described. At last, the joint video/depth rate control is proposed to adapt to the channel requirements. The main contributions of the thesis are summarized as follows.

Firstly, in multiview video coding, the view sequence coding at positions of the key frames contribute a great proportion in the whole coding performance of hierarchical-B pictures based multiview video coding. Based on this characteristic, a multiview image alignment method is proposed to improve the inter-view prediction efficiency. According to the multiview geometry, the parameters of warping alignment model are first estimated by global motion estimation, and then the reference images of adjacent views are warped to the coding image in the current view. Multiview image inpaintings are used to fill the blank holes in the warped images. The inpainted images are fed into the coding buffer as prediction reference images. To further improve the prediction efficiency of the warping aligned images, multi-hypothesis disparity predictions are used to build the jointly optimized weighting prediction signals. The experimental results illustrate that the proposed method can promote the view sequence coding performance efficiently.

Secondly, rapidly random access to some data section in the compressed stream can

guarantee the interactive feature in the video application. In current multiview video coding, more dependencies between frames occurred after introducing the inter-view prediction, and the random access to the compressed stream is correspondingly becoming very difficult. To solve the trade-off between the coding efficiency and random access, the corresponding methods are proposed to improve the performances of different kinds of random access. For temporal random access, a flexible view-temporal prediction structure regulating algorithm is proposed to improve the temporal random access ability. According to the random access constraint, the proposed algorithm can find the optimal view-temporal prediction structure to guarantee the optimal trade-off between the coding efficiency and temporal random access ability. For the view random access, the method of regulating inter-view prediction structure and the method of secondary representation coding are proposed to improve view random access ability. For view dimension coding, this thesis proposes three coding methods: interleaved view coding, view dimension coding using SP/SI frames, and secondary representation coding. These coding methods can improve the view random access ability while keeping higher coding performance. For the spatial random access, a novel multi-view video coding method is proposed to support spatial random access functionality. In this method, an independent space-time cube coding is used to limit the scope of inter-frame prediction, therefore random access of specific regions in the compressed data of a picture can be accelerated.

Thirdly, in multiview video system, the virtual view is generally synthesized by the compressed video and depth. The compression has a great effect on the virtual view synthesis. The effect of compression of texture and depth on the view synthesis quality is quantitative analyzed in this thesis. The depth error introduced by depth compression can cause the pixel position warping errors, and the compression-induced video intensity errors will transferred to the virtual view. Taking regard the virtual view synthesis as a linear view interpolation problem in mathematics, an upper bound of compression-induced view synthesis distortion is derived. The derived distortion upper bound depends on the depth-loss-induced pixel warping error and video-compression-induced intensity loss. Experimental results show that the proposed view synthesis distortion bound can approximately reflect the virtual view quality.

Fourthly, in multiview video and depth coding, the rate allocation between video and depth has a great effect on the virtual view quality. In this thesis, a view synthesis distortion model is proposed for joint video/depth rate allocation after investigating the effect of compression on view synthesis. The proposed distortion model is an additive model which accounts for the video-coding-induced distortion and the depth-quantization-induced distortion, as well as the inherent geometry distortion. Depth-quantization-induced distortion not only considers the warping error distortion, which is described by a piecewise linear model with the

video power spectral property, but also takes into account the warping error correlation distortion between two sources reference views. Geometry distortion is approximated from that of the adjacent view synthesis. According to the proposed distortion model, the synthesized view can be efficiently evaluated without access to the captured original view. Based on the proposed distortion model, a joint rate allocation method is proposed to seek the optimal trade-off between video bit-rate and depth bit-rate for maximizing the view synthesis quality. Experimental results show that the proposed distortion model is capable of approximately estimating the actual distortion for the synthesized view, and that the proposed rate allocation method can almost achieve the identical rate allocation performance as the full-search method at less computational cost. Moreover, the proposed rate allocation method consumes less computational cost than the hierarchical-search method at high bit-rates while providing almost the equivalent rate allocation performance.

Fifthly, how to provide a reasonable source coding rate to adapt to the channel bandwidth is a rate control problem which must be conquered in 3D video application. The existing rate control algorithms mainly focus on the 2D video coding. To provide a constant bit-rate control algorithm for 3D video services, this thesis first utilizes an image-stitching method to simultaneously encode video and depth, and then proposes a joint rate control algorithm for MVD. The proposed algorithm consists of three levels to achieve the accurate rate control. In the view level, different proportions of rates are allocated for different types of views. In the video/depth level, the target rates for video and depth are discriminatorily assigned to guarantee the high quality of video for the backward-compatible display. In the frame level, hierarchical rate allocation is used to regulate the target bits for each frame. In addition to the above mentioned rate control strategies, according to the special characteristics of multiview hypothetical reference decoder (HRD), the buffer-related rate control is also considered to prevent the decoder buffer from overflow or underflow even outputting multiple views. Experimental results show that the proposed rate control technique can accurately control the bit-rate to satisfy the requirement of 3D video systems.

Keywords: Multiview video, multiview depth, video coding, random access, joint video/depth rate allocation, joint video/depth rate control