近些年来，5G 和多媒体技术的发展引发了新一轮影像技术变革，从传统图像/视频加速进入以虚拟现实、增强现实以及混合现实为代表的沉浸式影像时代。点云是沉浸式视频的数据表达方式之一，被认为是走向沉浸式真三维表示和渲染的可靠途径，也是近年多媒体技术领域的研究热点之一。目前已经有很多三维点云的应用案例，包括沉浸式远程呈现、自动驾驶和游戏等。

点云由大量的离散空间点组成，每一个点包含了位置信息（X,Y,Z）和其他属性信息（颜色，法向量等）。其具有高维度且不规则的特点，点与点之间没有先后顺序，不存在拓扑结构信息。点云数据的特殊性导致目前已有的传统视频编解码器无法直接压缩点云数据。本文将探索点云数据组织方式与点云信号空域和频域分布特性，导出点云紧凑表示方法。同时，本文将基于点云稀疏离散分布，构建三维空间模型，提取三维空域特征与时域特征，并结合最优信号重建理论与率失真理论，实现点云的高效编码。本文的主要创新点如下：

针对点云空域建模与帧内编码问题，提出了一种基于主方向权重图的傅里叶变换方法。首先，基于位置信息构建空域图结构模型，通过邻域主方向和欧氏距离推导得到模型中的图边权重。利用图边权重表示空域相关性并导出图傅里叶变换，去除点云属性信息空域冗余；然后，提出了基于位置信息的定点 K 均值聚类算法，划分生成紧凑均匀的子点云，提升图傅里叶变换效率。

针对点云时域空域结构联合建模与帧间编码问题，提出了一种基于广义图傅里叶变换的帧间预测编码方法。首先，在点云空域图结构基础上，引入帧间参考关系（点云时域特征），并推导得到高斯马尔可夫随机场模型下的最优帧间预测和预测变换方法，去除点云属性信息时空域冗余；然后，提出了一种改进运动估计方法，解决了不规则点云相邻帧间时域对应关系搜索的难题；最后，提出了针对三维动态点云的编码框架，设计率失真优化模式决策方法，自适应选择最优编码模式，进一步提升点云属性帧间编码性能。

针对点云无拓扑结构空间建模与局部特征数据驱动建模问题，提出了一种多尺度端到端神经网络编码方法。首先，利用 Point Transformer 算子，生成嵌入位置信息和空域相关性信息的高维特征；然后，考虑到点云的稀疏性，设计多尺度邻域嵌入网络，学习点云的多层次细节特征，并生成点云高维紧凑表征；其次，在解码端，设计图空间扩展网络，恢复出邻域信息，提升重构点云质量；最后，采用最大倒角距离来衡量训练失真，有效缓解训练过程中出现的单边过拟合现象。相比于目前传统的位置信息编码方法，实现了约 34%的性能提升。

Recently, the development of 5G and multimedia technology leads to the advancement from traditional image/video to the era of immersive media, which is represented by virtual reality, augmented reality and mixed reality. As one of the data representation of immersive media, point cloud is regarded as a reliable approach to achieve immersive true 3D representation and rendering, and is also one of the research hotspots in the field of multimedia technology in recent years. There are many user cases of 3D point cloud, including immersive telepresence, autonomous driving and gaming, etc.

A point cloud consists of a large number of discrete spatial points, each point contains geometry information (X,Y,Z) and other attribute information (colors, normals, etc.). It has the characteristics of high dimension and irregularity, and there is no order among points and no topological information. Due to the particularity of point clouds, the existing traditional video codecs cannot directly compress point clouds. The thesis will explore the organization of point clouds and the spatial and frequency distribution characteristics of point cloud signals, deriving the compact representation of point clouds. Meanwhile, based on the sparse and discrete distribution of point clouds, a 3D spatial model will be constructed, and 3D spatial and temporal features will be extracted. The efficient coding for point clouds will be realized by combining the optimal signal reconstruction theory and rate distortion theory. Specifically, the major contributions of the thesis lie in the following aspects:

Aiming at the problem of point cloud spatial modeling and intra-coding, the thesis proposes the Normal Weighted Graph Fourier Transform (NWGFT). Firstly, the spatial graph structure model is constructed based on the geometry information, in which the graph edge weights are derived from the neighborhood normal vectors and the Euclidean distance. The NWGFT is derived from the graph edge weights which represent the spatial correlation, removing the spatial redundancy of point cloud attributes; Secondly, based on the geometry, the fixed-point K-means algorithm is proposed to generate compact and uniform sub-clouds, improving the efficiency of graph Fourier transform.

Aiming at the problem of point cloud spatio-temporal joint modeling and inter-coding, the thesis proposes a inter-prediction coding method based on the generalized graph Fourier transform. Firstly, based on the spatial graph structure of point clouds, the inter reference relation (temporal features of point cloud) is introduced to derive the optimal inter prediction and prediction transform under the Gaussian Markov Random Field (GMRF) model, removing the spatio-temporal redundancy of point cloud attributes; Secondly, the refined motion estimation is proposed to address the challenge of searching the temporal correspondence between neighboring frames of point clouds with irregular surfaces; Finally, a coding framework is presented for 3D dynamic point clouds, selecting an optimal coding mode adaptively, further improving the performance of inter-coding of point cloud attributes.

Aiming at the spatial modeling on non-topology structure of point clouds and the data-driven modeling of local features, the thesis proposes a multi-scale end-to-end coding method based on neural network. Firstly, the Point Transformer is used to generate high-dimensional features with embedded geometry information and spatial correlation information; Secondly, considering the sparsity of point clouds, the multi-scale neighbor embedding strategy is proposed to extract the features within level of details, generating the high-dimensional compact latent representation of point clouds; Then, in the decoder, the graph spatial extension network is proposed to restore the local spatial information and improve the quality of reconstructed point clouds; Finally, the max Chamfer Distance is adopted to calculate the training distortion, mitigating the unilateral over-fitting effectively during training process. It achieves about 34% performance increase, compared with the traditional geometry coding methods.