# 第二章 全景视频传输关键技术

随着多媒体技术和移动终端技术的快速发展，人们对全景视频业务的需求也与日俱增。全景视频有着不同于传统视频的新特性，因此如何在现有网络中传输全景视频涉及到多方面的理论与技术。本章首先介绍全景视频从采集到编码的基础知识，然后重点介绍了全景视频传输过程中的两个关键技术：视口自适应和无线资源分配，为后续改进算法提供理论依据。

2.1 全景视频基础知识

全景视频包含了360度的画面信息，用户可以转动头部来观看任意角度的视频画面。区别于传统视频的制作过程，全景视频的生成主要包含了采集、拼接、投影和编码四个过程。

2.1.1 全景视频采集拼接

现阶段主流媒体平台上的全景视频内容，通常是由多个不同角度的普通摄像机或鱼眼摄像机同时拍摄后拼接而成。目前市面上常见的全景摄像机如Google Jump、Insta360和GoProOmni和等均可采集较高质量的视频内容。

采集到多个角度的视频内容后需要进行拼接才可得到全景视频，拼接效果的好坏主要取决于图像配准的准确程度。现阶段主流的图像配准算法有基于全局灰度和特征点提取两种，其中基于全局灰度的算法容易受光线、噪声、相机定位等因素影响导致配准精度下降，而特征点提取算法如SURF（Speeded Up Robust Feature，加速鲁棒特征）、SIFT（Scale-invariant feature transform，尺度不变特征转换）、ORB（Oriented FAST and Rotated BRIEF，定向快速旋转二值）则更加稳定准确，其中SIFT检测法最为流行且技术成熟。

SIFT[53]算法于1999年由David Lowe提出并不断改进发展，改算法基于图像的局部关键点（特征点），不受图像缩放、旋转、平移及光照亮度的影响。SIFT算法主要有检测尺度空间极值、定位特征点、确定特征点方向和描述特征点几个步骤。

2.1.2 投影和编码技术

由于目前的视频编码和存储技术不能直接对三维的全景视频进行操作，所以首先需要将三维的全景视频映射到二维平面中，其中涉及到即视频投影技术。全景视频投影需要在投影后保留全部视频内容的同时防止过度扭曲变形的问题。现阶段常见的投影格式有ERP（Equirectangular Projection, 等角投影）和CMP（Cube Map Projection, 立方体投影）等。其中ERP是国内外主流内容平台都使用的一种投影方式，如图2-1所示。以地球仪展开为地图为例，ERP就是假设将地球仪嵌入一个圆柱体内，然后在地球仪的球心处放置一个灯泡，灯打开后地球仪上的画面照应到圆柱上，再把圆柱水平展开为二维平面即地图。这种投影方式在赤道部分的误差较小，越往两级误差越大。

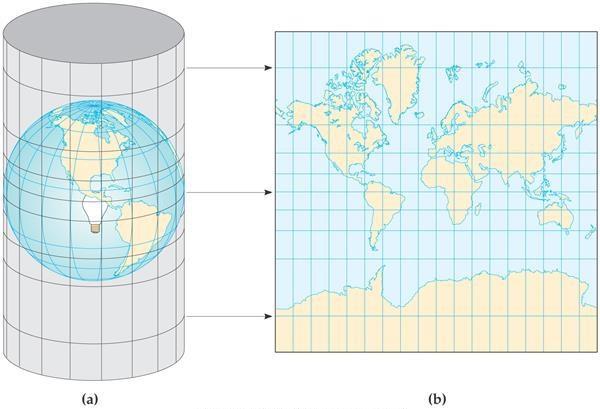


图2-1 ERP投影示意图

由于近几年全景视频才逐渐引起广泛的关注，目前业界内还没有制定出直接面向全景视频的编码技术，现有主流的商业方案是先将全景视频投影为平面视频后使用传统的视频编码技术H.264/AVC（Advanced Video Coding， 高级视频编码）或较新的H265/HEVC（High Efficiency Video Coding, 高效视频编码），后者有着更高的压缩效率和更复杂的计算过程。

考虑到现有编码技术处理全景视频时所产生的问题，学术界开始研究下一代编码方案。JVET（Joint Video Experts Term, 联合视频专家组）提出了新一代的视频编码标准H.266/VVC（Versatile Video Coding， 多功能视频编码），主要针对新兴应用如全景视频和高清视频，改进HEVC以提供更高的压缩性能，目前提出的方案理论上可以提高40%的编码效率。

2.2.3 视口自适应传输

目前全景视频传输主流方案有全部传输和基于用户视口传输两种方案，后者

通过视口预测算法预测出用户视口位置后，使用自适应流媒体技术有差别地传输视口和非视口区域的数据，可以很好地降低带宽需。自适应流媒体技术是一种主要基于HTTP协议的新型的流媒体技术，其中最为流行的是DASH（Dynamic Adaptive Streaming over HTTP，基于HTTP的动态自适应流媒体）。

DASH通过码率自适应的方式将不同质量的视频数据传输给客户端。首先将原始视频切分为若干个视频段序列，每个视频段被编码成多种码率版本存储在服务器上。然后根据网络状态和用户视口位置，客户端不断调整请求的码率版本，如当网络状态较好时，视口区域与非视口区域的视频段均选择较高码率的版本，而当网络状态较差时，优先请求高码率版本的视口区域视频段，在保证用户视口区域可以有较高质量内容的同时尽量避免非视口区域出现黑块或明显质量下降等问题的出现。

文献[7]首先使用线性回归预测用户视口位置，然后在带宽资源充足的情况下尽可能地提高视口内容的质量。由于该算法只传输视口区域，当预测出现误差时会导致视口区域出现黑块，严重降低了用户的观看体验。为了提高传输方案的容错性，文献[54]首先将全景视频源下采样得到一个低质量全视口版本，再结合高质量版本的视口区域内容一同传输给客户端，在降低传输数据量的同时保证了当视口预测失败时用户可以观看低质量的视频内容而不是出现黑块现象。文献[55]使用可伸缩编码将全景视频段编码成基础层和若干个不同质量的增强层，基础层保证最低质量的视频内容，然后根据信道条件和用户视口自适应选择增强层。这些自适应传输方案虽然在一定程度上降低了全景视频业务的带宽消耗，但未充分考虑视口预测失败的可能性，只通过传输低质量的非视口区域来保证观看体验，也造成了一部分的带宽浪费。