基于视口预测的全景视频传输关键技术研究

摘要

随着计算机科学与多媒体技术的发展，虚拟现实因其沉浸式构想式特性得到越来越多的关注，全景视频作为虚拟现实技术重要的组成部分也被广泛地应用到娱乐、教育、医疗和军事等领域。然而全景视频高码率低时延的特点对现有传输网络带来了极大的挑战。因此，如何在有限的网络带宽中最大程度地满足多个用户的观看体验至关重要。本文主要从全景视频视口预测和物理资源调度两方面来优化全景视频的传输。

近年来，基于视口自适应的全景视频传输方案得到越来越多的研究，而该方案的关键前提就是视口位置的精准预测。我们在现有研究的基础上，提出了一种基于用户历史观看轨迹的视口预测算法，首先根据当前用户的历史视口数据使用长短期记忆模型初步预测出视口中心位置的经纬度坐标，为了进一步提高长期预测的准确度，我们结合其他用户的视口数据对初步预测结果进行调整。实验结果表明，我们提出的预测算法比起基准算法能够取得较高的预测准确度。

另一方面，如何科学地调度带宽资源满足不同类型业务的用户需求也是研究重点。我们分析了LTE网络下经典的资源调度算法，针对经典算法在实时业务调度方面的问题，本文提出了一种基于Q学习的下行调度算法，可以根据实时变化的网络状态在每个调度间隔自适应地选择不同的调度算法以获得奖励最大化。仿真实验表明我们提出的算法在吞吐量、丢包率和时延等指标上优于经典算法，尤其在调度视频业务方面。

关键词 全景视频 视口预测 LTE 资源调度

RESEARCH ON PANORAMIC VIDEO TRANSMISSION

BASED ON VIEWPORT PREDICTION

ABSRTCT

With the development of multimedia technology, panoramic video has received widespread attention for its immersive experience and has been applied to entertainment, education, medical and military fields. However, the characteristics of high-bit-rate and low-latency of panoramic video pose great challenges to the existing transmission network. Therefore, how to maximize the satisfaction of the user's viewing experience in a limited network bandwidth is critical. This paper mainly optimizes the transmission of panoramic video from two aspects: viewport prediction and resource scheduling.

In recent years, more and more researches have been made on the viewport adaptive panoramic video transmission scheme, and the key premise of this scheme is the accurate prediction of the viewport position. Based on the existing research, we propose a viewport prediction algorithm based on the historical viewing trajectory of the user. First, based on the historical viewport data of the current user, a long-term short-term memory model is used to initially predict the viewport coordinates, and then combined with other users. Viewport data to correct prediction data. Experimental results show that our proposed prediction algorithm can achieve higher prediction accuracy.

On the other hand, how to scientifically schedule bandwidth resources to meet the needs of users of different types of services is also a research focus. We analyze the classic resource scheduling algorithm in the LTE network, and propose a downlink learning algorithm based on Q learning, which can adaptively select different scheduling algorithms at each transmission interval according to the network state to maximize performance. Simulation experiments show that our proposed algorithm is superior to classic algorithms in terms of throughput, packet loss rate, and latency.

KEY WORDS: panoramic video, viewport prediction, LTE, resource scheduling

# 第一章 绪论

1.1 研究背景与意义

随着计算机科学和多媒体技术的快速发展，VR（Virtual Reality, 虚拟现实）技术受到了人们越来越多的关注，虚拟现实技术构建的虚拟环境与真实环境真假难辨，用户从传统的以视觉为主的局限体验扩展到包含视觉、听觉、触觉甚至嗅觉的综合感官体验，在虚拟环境中体验到彷佛身处在真实世界的感觉。

虚拟现实主要有三大特性：沉浸式（Immersion）、交互性（Interaction）和构想性（Imagination）。沉浸式是指当用户处在虚拟现实模拟的世界中时，听到的看到的体验到的感受与真实世界完全一致，这是衡量虚拟现实技术优劣的关键指标；交互性指用户可以与虚拟环境进行互动。区别于传统的只能观看的用户体验，虚拟现实技术允许用户通过各种各样的传感器与虚拟环境进行交互，比如用户可以在射击游戏中拿起枪支对目标进行射击，可以蹲起躲避敌人，且整个交互过程遵循各种物理学定律，与真实世界一致；构想性也被称作想象性，用户在虚拟世界中除了可以体验到真实世界的感受外，还可以自由丰富地创建出真实世界不可能发生的事情，激发用户的创新性。

虚拟现实的概念可以追溯到20世纪80年代初，发展至今已经广泛地应用到影视娱乐、设计领域、医疗教育和军事等领域。2018年全球范围内虚拟现实的产业规模已近千亿元人民币[1]，中国的虚拟现实市场规模也在持续扩大，工业和信息化部预计到2021年，我国虚拟现实市场规模将达到544.5亿元，同时工信部表示将支持虚拟现实制造业创新中心筹备建设，推动虚拟现实技术在制造、教育、文化等领域应用[2]。虚拟现实巨大的发展市场，使得人们不断对其进行探索与研究。全景视频作为虚拟现实技术最为流行的应用方向之一，也逐渐成为人们争相研究的重点。

全景视频，又被称作360度视频、沉浸式视频，是一种用户可以随意转变观看角度的三维视频。传统的二维视频的缺点主要是缺乏互动性，用户只能被动观看视频内容，无法像在真实世界中那样可以随意转动头部观看不同的场景画面，全景视频的出现则打破了这种僵局。目前在各大网络平台看到的全景视频主要是利用由多个摄像机组成的全景摄像机拍摄多个角度的视频画面，然后通过特征点提取、特征匹配和融合等一系列步骤将各个角度的视频画面拼接成360度得到的。用户在观看全景视频时需要佩戴特殊的硬件设备VR眼镜或者HTM（Head Mounted Display, 头戴式显示仪），市面上常见的有Facebook的Oculus 和 HTC的Vive，然后用户就处在全景摄像机的位置观看360度的视频画面，用户可以自由地转动头部，视频画面也会随着变化，如同在真实世界中一样，获得一种身临其境的视觉体验。

由于相对于一般的平面视频而言，全景视频更加贴近用户的眼睛，对视频分辨率有着更高的要求，所以即使在相同的单眼分辨率情况下，由于全景视频需要保留360度的画面信息，全景视频的码率也要比平面视频的码率高很多，通常为前者的5倍以上。同时，全景视频对网络系统的传输时延有着较为严格的要求，业界公认的传输系统的MTP（Motion-to-Photos Latency, 头动到视野延迟），即转动头部和视频画面对应切换之间的响应时间必须控制在20ms以内[3]，否则会引起观看者的眩晕感，造成极差的观看体验。因此，全景视频“高码率、低延时”的传输特性对现有的无线传输网络提出了极大的挑战，如何在无线网络中为多用户提供高质量的全景视频服务逐渐成为亟待解决的问题。而在实际观看场景中，由于人眼角度的特性，在某一时刻，用户实际只能看到全景视频中的部分内容，称之为FOV（Field Of Viewport, 视口区域），如果以同等质量传输所有的画面信息势必会造成网络资源的巨大浪费，如何根据视口区域自适应传输全景视频内容也收到越来越多的关注。

另一方面，随着终端技术的发展，接入无线网络的智能设备愈来愈多，无线网络传输的业务类型也多种多样，从传统的二维视频、音频及图文到新兴的全景视频等等，不同类型的业务对网络的QoS（Quality of Service, 服务质量）的需求不一样，如全景视频这种实时业务对系统的传输时延有着较高的要求，而文本这种非实时业务则要求传输系统有着较低的丢包率来保证信息的完整度，如何合理利用有限的带宽资源来满足不同业务的需求也是极为关键的问题之一。在LTE（Long Term Evolution, 长期演进）和5G（5th-Generation, 第五代移动通信技术）中，基站侧的无线资源管理起着极为重要的作用，而其中涉及的资源调度算法则是关键。所谓资源调度，就是通过一定方式计算小区内等待调度的业务的优先等级，然后根据等级顺序将带宽资源依次分配给业务，直到所有待调度业务都满足了传输需求或者带宽资源全被占用。通常而言，资源调度算法需要考虑网络时延、吞吐量、时延以及用户间的公平性等多个方面，一个高效的调度算法则可以在各个方面取得均衡，使得效益最大化。因此，如何有效地利用有限的系统带宽资源，通过合理的调度与分配，为多用户提供高质量低时延的全景视频业务，同时也保证其他传统业务的服务质量，也是目前全景视频传输优化工作的重点与难点。

1.2 国内外研究现状

全景视频一直以来都是虚拟现实领域的重点研究方向，涉及到的技术包括全景视频的拼接、投影到基于视口的自适应传输再到客户端的渲染重建。本文主要聚焦于全景视频的视口预测算法和底层物理资源调度两方面。

1.2.1 全景视频视口预测研究现状

目前全景视频的传输方案主要分为全部传输和自适应传输两种[4]。全部传输是类似于传统视频传输方案，即将全景视频完整的传输给用户，目前主流的流媒体平台（例如YouTube）都是采用这种传输方式。这种传输方案可以保证用户看到全部的视频内容但是对网络负担极大，且会造成较多的带宽浪费。自适应传输方案则是针对视口区域进行差别传输，其中基于tile（区块）的传输方式被认为是传输全景视频的理想方案[5][6]。不同于传输整个画面，基于tile的方式以较高质量传输视口区域所在的tile，而以低质量传输非视口区域的tile甚至选择不传输。这种传输方式可以很好的降低带宽需求，但很依赖准确的视口预测算法，一旦视口区域估计错误，将会导致用户观看区域内出现明显的低质量画面甚至是黑块区域，这将会严重降低用户的观看体验，因此，如何准确有效地预测出用户的视口区域对全景视频的自适应传输有着极为重要的作用。

对视口进行预测，首先要明确的问题是视口是否可预测。文献[7]对视口中心的位置，即视点（viewpoint）数据进行了实验分析。文献主要研究表征视点位置的三个参数：偏航角（yaw）、俯仰角（pitch）和翻滚角（roll）的相关特性。首先分析了三个角度的累积分布图，实验结果表明偏航角的变化幅度最大，最难预测，其次是俯仰角；然后又分析了三个角度和各自差值的自相关函数，结果显示在500ms内有较强的自相关性，这表明视点数据是可以预测的。

目前全景视频视口预测的方法主要有两种：基于运动轨迹的预测和基于内容检测的预测。前者一直是研究的重点。

基于轨迹预测主要是根据该用户历史一段时间内的视口位置来预测接下来的视口位置，视口位置主要通过头部移动轨迹或者眼球移动轨迹来确定，而这些轨迹数据可以通过HMD的定位功能采集到。

基于运动轨迹的预测可归类为时间序列预测问题，因此很多时间预测的方法被应用进来，如滑动平均、线性回归和机器学习等。文献[8]对比了滑动平均、线性回归和加权线性回归三种预测算法分别预测yaw、pitch和roll三个角度的准确度。滑动平均复杂度最低，即使用历史窗口中所有视点数据的采样点的平均值作为下一时刻的预测值，当数据变化幅度较小时这种预测方法也可以获得不错的预测准确度。线性回归则是利用统计学中的回归模型，根据历史一段时间内的用户视点数据进行拟合，然后通过拟合出来的函数模型获取下一时刻的预测值，加权线性回归则是给予历史数据不同的权重值，如越近的数据拥有越高的权重值，以此来获得不同的影响程度，一定程度上提升预测准确度。[8]的结果表明，线性回归与加权线性回归预测结果要好于滑动平均，当使用历史1s的视点数据预测接下来1s的视点数据时，加权线性回归的预测准确度高达90%以上，但是随着预测时间的增加，预测准确度将会明显下降，当预测时间增加到2s时，预测准确度则会下降到70%。文献[7]则对比了朴素预测，即使用当前的角度值当作预测值，线性回归和神经网络预测未来0.1s~1s范围内的偏航角，实验结果证明神经网络的预测效果要好于前两个算法。

除了使用当前用户的视点数据外，一些研究还探索了跨用户的行为。文献[9]使用了结合KNN（K-Nearest Neighbor, K最邻近） 聚合的线性回归模型，基于当前用户的历史视点，使用线性回归模型预测出当前用户下一时刻的视点后，再结合观看同一视频的其他用户的历史数据，使用KNN来对预测结果进行纠正，进一步提升预测准确度。文献[10]对三个角度的轨迹进行建模，具有相似观看行为的轨迹被聚合在一起，并且为每个聚类计算出不同的轨迹函数。文献[11]则首次使用一种基于密度的聚类算法DBSCAN，在服务器端先对用户进行分类，然后在客户端，通过SVM（Support Vector Machine, 支持向量机）分类器来预测用户的类别，最后预测该类别所对应的观看轨迹。

另外一类是基于内容检测的视点预测。该方法主要从视频显著性的角度进行预测。视频显著性表征了用户对视频不同内容的关注程度。一般来说，视频内容的显著性越强，用户越关注，则内容被观看的概率也越高。近年来，基于深度CNN（Convolutional Neural Networks, CNN）的显著性检测方法相比于传统方法可以取得较优的结果[12][13]。文献[26]训练了一个浅层网络进行显着性检测，并应用了转移学习生成一个深层网络；同样的作者在文献[27]中介绍了SalGAN，这是一种通过对抗进行训练的深度网络；但是，这些研究工作都是在常规的平面图像上进行的，由于从三维球体到矩形平面的投影中引入了较大的畸变，因此将这些技术直接迁移到全景图像上注定会失败。为了解决这个问题，最近的文献[28]提出构造依赖位置的过滤器，即内核的大小根据所应用的纬度而变化；文献[29]则是提出了一种用于微调全景图像的传统二维显着性预测网络的管道。将等角矩形图像分成六个等大小的块来渲染每个视口区域。基础网络得到了进一步的优化。

为了更准确地捕获未来视口与过去视口之间的非线性和长期依赖关系，文献[21]使用卷积神经网络开发了一个视口预测模型，模型删除了池化层，并添加了更多的卷积层以增强非线性拟合能力;文献[22]提出了两种深度增强学习模型：首先使用一个仅基于视觉特征来估计视频中每帧热图的离线模型，然后使用一个在线模型，根据过去观察到的头部移动位置以及离线模型得到的热图来预测头部运动。文献[23]提出了一个固视预测网络，该网络同时利用过去的视口位置和视频内容特征来预测接下来n帧中的视口轨迹或基于图块tile的观看概率。

在时间序列预测场景中常被使用的长短期记忆模型(Long-Short Term Memory, LSTM)也被逐渐应用在视口预测中。LSTM是一种改进之后的循环神经网络(Recurrent Neural Networks, RNN)，可以解决RNN无法处理长距离依赖的问题;文献[24]使用LSTM编码FoV扫描路径的历史记录，并将隐藏状态特征与视觉特征结合起来进行长达1秒的预测;文献[25] 提出了一种基于LSTM的深度神经网络（Deep Neural Networks , DNN）模型。该模型融合了头部移动轨迹、跨用户热图和视频显著性检测，可以共同预测用户未来的观看方向。文中在记录了100多个用户的观看的数据集上评估了所提出的方法，并表明提出的方案优于几种基准方案。

1.2.2 LTE资源调度算法研究现状

为了满足无线网络中多用户不同业务的服务质量需求，如何对有限的无线带宽资源进行有效的资源调度一直是学术人员的研究方向。

最初学者们提出了一些面向非实时业务的经典的调度算法，如轮询算法（Round Robin, RR）[1,2]，该算法将资源轮流公平地分配给网络中的用户，而不考虑实时变化的信道条件和系统吞吐量；最大载干比算法（Maximum Carrier to Interference, Max C/I）[3,4]，该算法则是以系统吞吐量最大化为目标，只考虑用户信道条件的好坏；比例公平算法(Proportional Fairness, PF)，该算法结合了前两个算法的特点，综合考虑了系统吞吐量和用户公平性，所以是无线网络中最常使用的调度算法。但是这几种算法都没有考虑不同类型业务的（Quality of Service, QoS）需求，在实时业务的场景中性能较差。鉴于视频类实时业务需求的增加，研究人员们又提出了一些面向实时业务的经典算法，比如最早到期优先算法(Earliest Deadline First, EDF)[5]，该算法根据每个业务的QoS需求中的最大时延对所有任务队列进行排序，每次优先调度截止时间最小的业务，保证了时延敏感业务的调度，但是忽略了系统的整体吞吐量和用户之间的公平性；改进的最大权重时延优先算法(Modified Largest Weighted Delay, M-LWDF)[6][7]，该算法基于比例公平算法，综合考虑了分组数据的丢弃时延和信息质量条件，可以在系统吞吐量和实时业务丢包率之间取得较好的平衡。

在这些经典的调度算法的基础之上，大量改进的优良算法也被陆陆续续地提出。基于PF算法，文献[8]对非实时业务和实时业务进行了区分，对于实时业务，改进的算法考虑了其时延特性，保证了时延较为敏感的业务可以优先被分配资源得以调度；文献[9]对两种业务的调度算法均进行了研究，提出了一种改进的时延优先级函数，实验结果表明，改进的算法在系统吞吐量、用户丢包率和公平性等指标均得到了提升；

除了从网络系统方面对算法作出优化外，研究人员也从用户侧出发，以用户体验(Quality of Experience, QoE)为优化目标改进调度算法。文献[9]根据QoE指标，即平均意见值得分(Mean Opinion Score, MOS)，评估了LTE网络中三种流行的下行链路调度算法(PF、EXP-PF和M-LWDF)。实验结果表明，最适合的下行调度算法是M-LWDF，其端到端时延小于50ms，并且在可以接受的MOS得分(超过3.5分)下运行最大数量的用户接入(超过50个用户)。当用户超过30时，最为广泛使用的PF算法的端到端时延会增加到200ms以上，这表明PF算法并不适合调度实时业务；文献[10]分析了视频传输流程与考虑重点，探索了LTE网络中传输视频的QoE改善问题；

为了更好地结合多方面的因素来达到高效调度的目的，跨层调度也得到了深入的研究。文献[11]提出了一种跨层方案，协同无线协议栈的应用层、MAC层和物理层参数共同优化，最大化利用网络资源并提升用户的感知服务质量。仿真结果显示提出的跨层框架可以最大程度的提高用户的感知质量并且在用户的公平性方面得到了显著的改进。文献[12]则在跨层优化的基础上，采取了多点协作的方式，进一步降低了实时业务的传输时延。

在最近的研究中，深度学习的相关方法也被应用在无线网络传输中。针对资源调度问题，文献[13][14]聚焦于基于深度Q学习(Q-Learning, QL)的调度算法，建立马尔可夫决策过程(Markov Decision Process, MDP)模型来代表系统中各个状态的改变，然后根据不同类别用户的信道质量指标(Channel Quality Index, CQI)在每个传输时间间隔采用不同的调度规则。

1.3 主要研究内容

随着全景视频应用的增长，如何优化全景视频的传输成为了一种重要的研究方向。针对全景视频的视口预，基于运动轨迹的预测没有考虑视频内容对用户观看行为的影响，在长期预测场景中存在较大问题；而基于视频显著性检测的算法复杂度不利用对时延比较严格要求的全景视频，同时显著性检测也未能考虑观看者在年龄、性别、兴趣爱好等方面的区别，因此在内容提取方面也存在一些误差。

随着全景视频等视频业务在无线网络中的增加，如何科学调度分配有限的带宽资源来满足不同用户不同业务的需求，也是亟待解决的问题之一。本文提出了一种新颖的调度算法，根据吞吐量、时延和公平性指标定义了若干种状态，然后基于Q学习算法来觉得在每种状态下应该选择何种调度策略来达到系统的性能最优化。

1.4 本文组织结构

本文共分为四个章节，各章节安排如下：

第一章：绪论。该章节主要介绍了全景视频传输的研究背景和研究意义，针对传输框架中的视口预测和资源调度两方面，简要阐述了相关的国内外研究现状和研究内容，最后给出本文的组织结构；

第二章：基于历史观看轨迹的视口预测算法。本章介绍了现有阶段视口预测的相关研究，分析了常见预测方案的不足之处，然后详细介绍了本章提出的基于历史观看轨迹的视口预测算法，首先根据当前用户的历史视口数据，使用长短期记忆模型初步预测未来的视口数据，然后再结合观看同一视频的其他用户的视口数据对预测数据进行矫正，得出最终的视口所在位置。最后阐述了相关实验过程与结果对比分析；

第三章：基于Q-Learning的LTE资源调度算法。本章首先分析了LTE网络的基础架构和关键技术，然后介绍了常见的资源调度算法与资源调度流程，在此基础上，本章提出一种基于Q-Learning算法的下线资源调度算法。该算法根据动态变化的网络状态，使用Q-Learning算法在每个传输间隔选择不同的调度算法，最后使得系统性能最优化。最后介绍了仿真平台与仿真环境，从系统吞吐量、丢包率和用户业务时延等方面对实验结果进行了分析与对比。

第四章：总结与展望。本章节对论文做了回顾总结与未来展望。

# 第二章 基于历史观看轨迹的视口预测算法

根据第一章的介绍，现阶段国内外视口预测主要分为两种方式：基于用户历史观看轨迹和基于视频内容。由于基于视频内容的视口预测算法时间复杂度较高，不适合全景视频直播场景，同时，不同年龄、性别的用户在观看视频时的感兴趣区域差异较大，通过视频敏感性来预测视口会造成较低的预测准确度，因此，本章只聚焦于基于历史观看轨迹的预测算法。首先介绍视口预测相关技术的理论基础，然后详细介绍基于神经网络的预测算法，并结合其他用户的视口数据辅助预测结果，最后，对比分析了其他算法与我们算法的预测准确度。

2.1 视口预测相关技术

视口预测作为全景视频传输优化问题的前提之一，已经出现了一些针对该问题的研究。从本质上来说，视口预测可以归为时间序列预测问题，本节主要介绍了视口的相关定义和常见的时间序列预测算法。

2.1.1 相关定义

（1）视口区域

在给出视口区域的定义前，首先介绍人眼特性和VR眼镜（或头显）的特性。

人类眼睛单眼在水平方向最大可以观看到150度左右的画面，双眼在水平方向则可高达188度左右。但是双眼重合视区为120度左右，即在这个角度范围内，观看到的物体才会呈现出立体感。超过水平视角30度之外的视角称为诱导视角，即我们常说的余光，在这些区域的事物人眼是不敏感的。而在垂直方向上，单眼感光区域约为120度，视觉敏感区为60度，如图2-1所示。

### (a) 水平视角 (b) 垂直视角

### 图2-1 人眼视角区域

由于全景视频的特性，必须佩戴相关的观看设备（如VR眼镜或者头戴式显示仪）才能观看。这些硬件设备在生产时都会有各自固定的视角区域。目前市面上入门级的vr眼镜的视角区域水平为90度，垂直区域为90度；像Facebook的Oculus和HTC Vive之类昂贵的头显设备的视角区域则为水平110度，垂直90度。

全景视频是一个横向呈360度，纵向成180度的三维视频，但由于人眼特性和全景视频观看设备的参数，在某个特定时刻，用户只能观看到全景视频的一部分内容，这部分就称为视口区域（Field Of Viewport, FOV），如图2-2所示：



### 图2-2 视口区域

为了方便后续的实验说明，本文中所指的视口区域特定为水平110度，垂直90度的矩形区域。

（2）视点

视口区域是一个矩形区域，我们定义该矩形的中心点所在位置为视点。当视口大小固定时，只要知道视点位置即可获取到视口区域。我们给出可以描述视点的几种常用参数如下：

a） x、y、z坐标

我们将全景视频构建成笛卡儿坐标系（Cartesian Coordinates）中的一个单位圆，视点位置即为圆上一点，可以使用空间三维坐标x、y、z来表示

b）经纬度

全景视频可以通过多种投影方式展开为二维平面视频，常见的投影方式有等距柱状投影（Equirectangular Projection, EPR）、多面体投影（Platonic Solid Projection, PSP）和立方体投影（Cube Map Projection, CMP）等。我们使用国内外主流内容平台都使用的EPR方式将全景视频投影为二维平面视频，如图2-3所示。ERP投影简单来说，就是经典地将三维地球仪打开成二维地图的方式。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2-3 EPR投影示意图

经投影展开后，三维球面上的视点就落在了平面上，我们使用经纬度来表示。经纬度与三维坐标可以通过下面的公式互相转化：

(2-1)

其中，R代表球体半径，为维度，为经度。

c）偏航角、俯仰角和翻滚角

用户在佩戴vr设备观看全景视频时，头部从某个位置转到另外一个位置的旋转量可以使用三个欧拉角（Euler Angles）来表示，即偏航角（yaw）、俯仰角（pitch）和翻滚角（roll），如图所示。俯仰角为头部旋转的水平夹角，向右转正方向，范围为-180度到180度；俯仰角为头部旋转的垂直夹角，向上抬头为正方向，范围为-90度到90度；翻滚角则是视线看向正前方，头顶在视线的正交平面左右摆动，以向右摆为正方向，范围为-90度到90度。

图 方向角示意图

d）四元数

旋转量使用欧拉角表示直观又简单，但是这种表示方式存在万向节死锁的问题。万向节死锁简单来说，就是如果使用欧拉角来表示旋转过程的话，在某些特定位置，三个互相正交的平面会出现其中两个平面重叠的现象，导致三维信息丢失。为了解决这个问题，研究人员提出了四元数的概念。四元数基于一个定理，即三维空间中的一个物体，绕着某个特定的方向向量，旋转特定的角度，可以到达三维空间中的任意一个位置。四元数本质来说是一个高阶负数。四元数与x、y、z可以通过如下公式进行转化：

公式 四元数转欧拉角

解释一下公式含义。

（3）视频区块

现阶段，全景视频的传输主要有两种方式，一种是传输全部内容，一种是传输部分内容。第二种传输方案中，需要从空间上将全景视频分割成不同的区域，称为区块（tile），如图2-5所示，我们将全景视频分成了8\*4个区块。假设视口区域为A，则基于tile的传输方式只需要传输视口区域所覆盖的6块tile，如图中黄色区域所示。



### 图2-5 tile示意图

2.1.2 常用预测模型

全景视频视口预测主要有两种方法，基于历史运动轨迹和基于视频显著性检测。考虑到后一种方法的时间复杂度不太适合实时预测，本文只探讨第一种预测算法。

基于历史运动轨迹的预测，通常是使用历史的视点数据来预测接下来一段时间的视点数据，本质上可以划分为时间序列预测问题。时间序列的预测是数据挖掘领域热门的研究方向，研究人员们已经提出了许多优秀的预测模型，主要可以分为基于统计学的，基于机器学习算法的和基于神经网络的，本节介绍这三种类别中的具有代表性的预测模型。

（1）移动平均

移动平均（Moving Average, MA）是一种简单的时间预测模型，主要思想是根据历史一段时间（即窗口）内序列的平均值来预测接下来某一时刻的序列值，然后移动窗口继续预测，计算公式如下：

其中，表示t时刻的预测值，为历史n个序列值。

当数据变化范围较小，整体呈现平稳状态时，该方法可以取得较好的预测准确度。在移动平均预测的基础上，又提出了加权移动平均法，即给历史窗口中的每个序列不同的权重值，计算公式如下：

其中，为第n个权重值。

使用该方法进行预测时，权重值的选择尤为重要，最为简单的是经验法和试验法，通常情况下，最近的序列值最能影响未来预测值，所以权重值应该大一些。

（2）ARIMA

整合移动平均自回归模型（Auto-Regressive Integrated Moving Average , ARIMA）是基于统计学的十分流行的一种预测模型。该模型有着完善的理论依据，简单高效。

ARIMA模型是在移动平均自回归模型（Auto Regression and Moving Average, ARMA）的基础上进行d阶差分得到的。差分是将不平稳序列转为平稳序列的常用方法之一，所谓差分，就是使用后一个序列值减去前一个序列值后的结果，一阶差分即进行一次减法运算，二阶差分则在一阶差分的结果上再进行一次减法运算，d阶差分依此类推，计算公式如下：

ARMA模型则是由自回归模型（Auto Regression, AR）和移动平均模型（Moving Average, MA）模型组成，表达公式如下：

其中，是模型的输出，即t时刻的预测值，是AR模型的公式，是系数，代表时间序列值；AR模型主要是探索预测值与历史序列值之间的线性关系。而MA模型主要是探索预测值与预测误差的线性关系。由公式可以看出，ARIMA模型的关键是确定三个超算数p, q, d的值，差分中的d可以通过平稳性检验来确定，p和q一般是通过观察时间预测自相关图（Auto Correlation Function, ACF）和偏自相关图（Partial Auto Correlation Function, PACF）[39]的截尾和拖尾现象来确定。

ARIMA模型简单高效，但仅适用于线性规律，同时关键的p和q的值需要人工判断，且参数一单固定后就无法应对数据的变动，在实际的视口预测模型中，由于用户观看不同视频时的观看行为不同，需要对每一个视频都进行参数估计建立新的模型，工程量大效率低下，因此该模型并不适用于本文探讨的预测场景。

（3）线性回归

随着机器学习技术的发展，研究人员也开始使用一些机器学习的模型来进行预测。线性回归（Linear Regression, LR）[40]是机器学习中较为基础的一种算法。线性回归主要是找到数据之间的规律函数，对于单变量而言，线性回归拟合出一条直线，方程可表示为：，对于多个变量的话，拟合出来的表达式可表示为：

其中，是常数项，为对应的偏回归系数。

基于机器学习的预测算法主要是模型的训练，一般是根据历史窗口的大小将数据集划分为多个子序列段，每个子序列段后的h个值就是待预测的真实值，模型输出的即为对应的h个预测值。在这种方法中通常使用真实值与预测值的均方误差（Mean Squared Error, MSE）做为损失函数（Loss Function），计算公式如下：

其中，n为预测值个数，为系数，为真实值，为对应的预测值。

回归最后就转变成了优化问题，即找到使损失函数最小的解，划分好训练集与测试集之后训练模型即可。

除了线性回归模型，常用的机器学习模型还有支持向量机模型（Support Vector Machine, SVM）[41]和随机森林模型（Random Forest, RF）[42]。相比于基于统计学的预测方法，基于机器学习的方法可以处理非线性关系，同时训练出来的预测模型具有普适性。

（4）多层感知机

由于深度学习的崛起，一些神经网络的模型也被应用到时间序列的预测中来。与机器学习算法类似，我们也可以通过训练神经网络模型来获取到历史序列值与预测值之间的规律关系。

多层感知机（Multilayer Perception, MLP）[47]是最基本的人工神经网络模型，而最经典的MLP分为输入层、隐藏层和输出层三层网络，不同层之间使用全连接，如图所示。

### 图 MLP示意图

隐藏层的输出是由三部分构成的：权重、偏置和激活函数。其中，权重代表着神经元之间的连接强度，权重越大，可能性越大；偏置是模型中重要的参数，偏置的设置是为了正确分类样本，保证通过输入算出的输出值不能随便激活；激活函数起到非线性映射的作用，可将神经元的输出幅度限制在一定范围内，一般限制在-1～1或者0～1之间。最常用的激活函有Sigmoid函数、tanh函数和ReLU函数等，采用哪种激活函数得视具体情况而定。

MLP被应用到了许多预测场景中。文献[43][44]对比了ARIMA模型和MLP模型在时间序列预测方面的性能差异，实验结果表明，短时间内MLP可以获得更高的预测准确度。文献[45]使用MLP预测了河流流量，文献[46]则使用MLP预测了金融领域的数据。

（5）循环神经网络

循环神经网络（Recurrent Neural Network, RNN）[48]是以序列为输入，针对序列数据进行建模的深度神经网络。基础的神经网络只在层与层之间建立了全连接，RNN最大的不同之处在于层之间的神经元也建立了全连接。RNN神经网络的结构图如下：



图 RNN结构图

相比起基于统计和基于机器学习的模型，神经网络拥有强大的拟合能力，尤其是针对非线性关系的序列数据，文献[49]证明一般的神经网络可以任意逼近任何非线性函数。当前主流的神经网络都是基于梯度反向传播算法进行训练，可以实现端到端的训练过程，且可以增量式更新模型权重，有利于模型应对多变的环境。神经网络具有多种连接形式，以RNN为代表的模型可以针对序列数据的前后依赖关系建模，可以更好地处理时间序列相关性的问题。

2.2 基于历史观看轨迹的视口预测算法

根据上述介绍可以看出，针对视口数据这种呈现非线性关系的时间序列进行预测，使用深度神经网络更能取得较好的预测准确度。目前的研究文献中主要提出了基于线性回归和神经网络模型的预测算法预测准确度随着预测时间的增长下降明显，无法做到长期预测。

为解决上述问题，本文提出了一种基于历史观看轨迹的视口预测算法，首先使用LSTM模型（Long-Short Term Memory, 长短期记忆）分别预测当前用户视口中心位置的经度和纬度坐标，然后结合其他用户观看同一个视频时的观看轨迹，计算出最终视口区域所落在的tile位置。

2.2.1 数据集介绍

视点数据集的合理选择是预测算法成功的重要前提。由于视频内容、用户特性对用户的观看轨迹影响很大，因此合理的数据集应该包含各种类型的全景视频，观看视频的用户也应该分散到各个群体。

我们分析了现有公开的几个视点数据集[52][53][54]，最终使用wu等人采集的全景视频用户头部移动数据集[54]。该数据集为48个用户观看18个全景视频的头部移动数据，48名用户包括24名女性和24名男性，覆盖了各个年龄范围，18个全景视频则包含了演出、体育、电影、脱口秀和纪录片5种最流行的类别。

18个全景视频的时长从2分44秒到10分55秒不等，帧率为30fps，而头部移动采集设备每秒可以采集60~80个数据，且每两个采样点之间的时间间隔有轻微的随机性，为了后续数据分析的便捷性，本文使用线性插值法将原始数据下采样为每秒30个数据，与视频的帧率保持一致。最后我们将原始数据转换为经度和纬度值，其中经度的取值范围为0~360度，纬度的取值范围为0~90度。

2.2.2 LSTM介绍

由于后续的视口预测算法是基于LSTM模型的，本节先介绍一下LSTM的相关理论知识。

RNN在捕捉数据在长时间范围内的关系方面存在着明显的缺陷。使用递归结构，隐藏层到隐藏层到转换矩阵（公式 ，在介绍RNN时给出）被多次使用。例如，为了基于第一次输入来更新输入权重U，误差值需要通过整个路径向后传播到第一个时间步长。如果W的最大特征值大于1，则会造成梯度爆炸问题，而如果W梯度到最大特征值小于1，则会造成梯度消失问题，在训练中起不到任何实际作用。梯度爆炸的问题可以通过梯度截断来解决，但梯度消失的问题很难解决。

为了对更长时间的依赖关系进行建模以及解决RNN梯度消失的问题，长短期记忆（Long Short -Term Memory, LSTM）[50]被提出。LSTM引入了门函数来控制信息流。与传统的RNN相比，LSTM的关键特征是使用ct表示的存储单元，它用来连接时间序列并且充当累加器。该单元可以通过几个参数化的控制门（输入门、忘记门和输出门）来进行访问、写入和清除。输入门控制当前输入xt和最后一个隐藏状态ht1。如果忘记门ft开启，存储在单元中的信息也可以被丢

弃，即忘记。最后，LSTM的输出是由输出门ot控制并通过tanh函数驱动的的，输出值介于-1和1之间。LSTM的基本单元结构表达式如下：

公式 LSTM表达式

2.2.3 视口预测模型

由前面的分析可知，当视点位置确定时，视口区域也可以确定，因此，本节聚焦于视点的预测，根据历史的经纬度数据来预测未来一段时间内的经纬度数据。

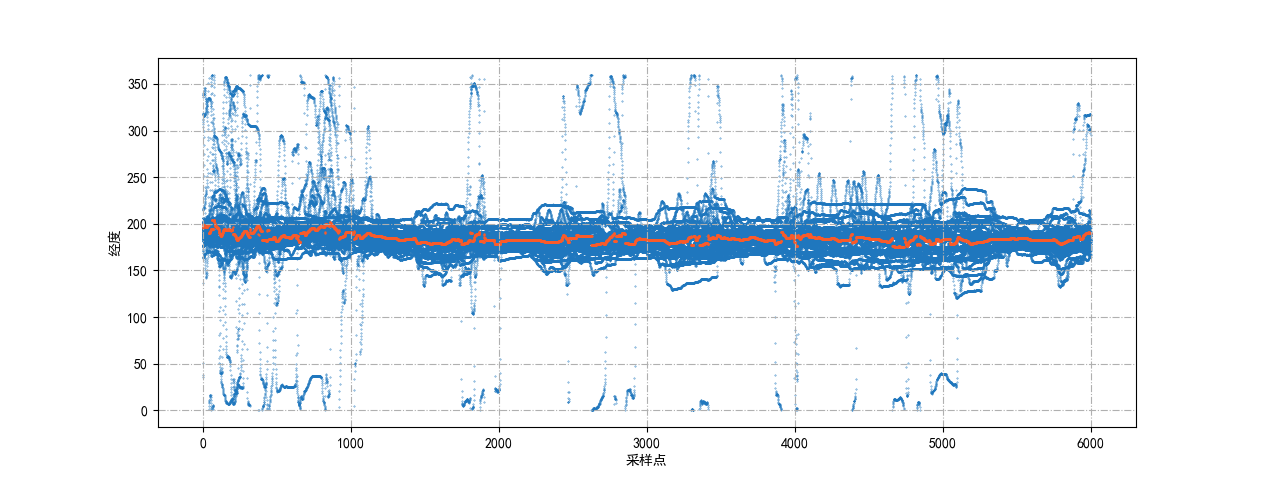
假设和分别为t时刻的经度和维度值，和为用户初始的经纬度，和分别为到时刻之间的经纬度值，即和。在t时刻的时候，用户的头显设备记录下来从初始值到t时刻的全部的经纬度值（），视点预测即使用历史窗口中的若干经纬度值来预测接下来个时刻的经纬度值，记为（）,其中为预测窗口，然后移动历史窗口继续预测，直至视频观看结束。文献[]认为，视点的几个特征维度之间的自相关性大于互相关性，因此本文也采用独立预测经纬度的方式。考虑到经纬度序列的时间特性和非线性特性，我们引入LSTM模型进行长期时间的预测，历史窗口的数据即为模型的输入，而模型的输出则为我们想要的经纬度的预测值，模型结构图如下所示：

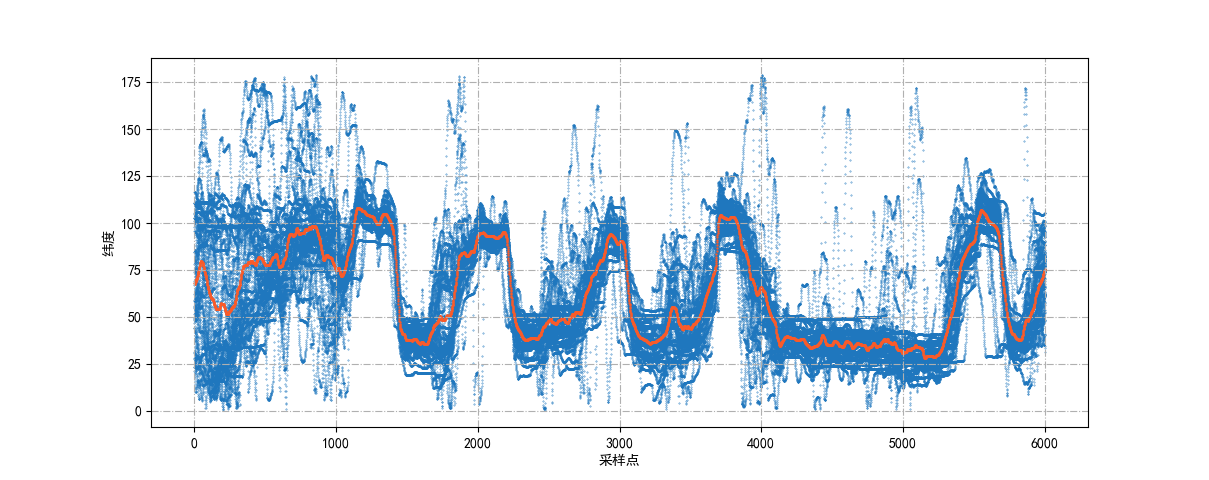


图 基于LSTM的预测模型结构图

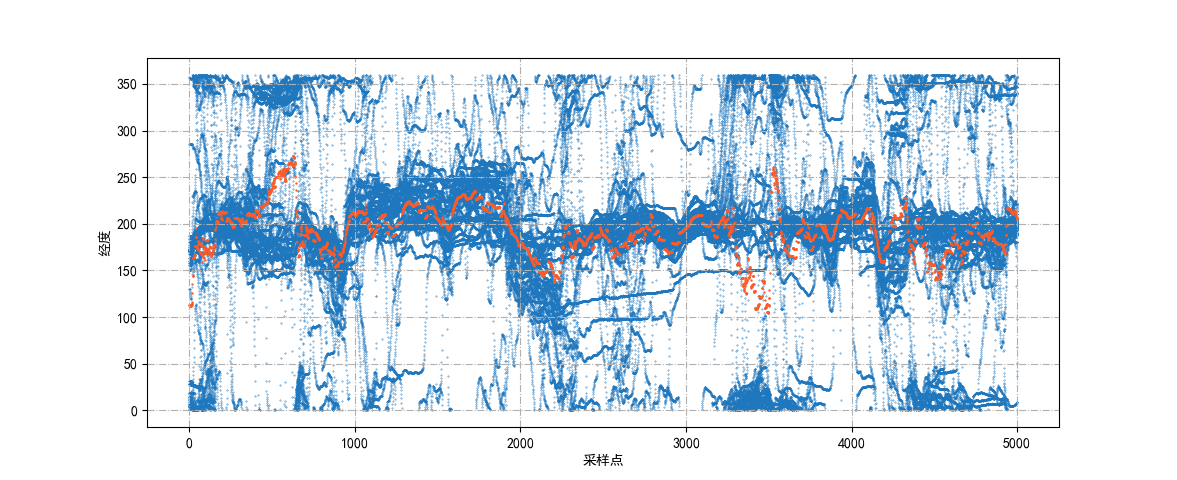
从图中可以看出，在t时刻可以根据历史的一段数据预测出下一个时刻的数据，表示为，其中为训练好的LSTM模型的输入输出函数；为了继续预测接下来的数据值，我们将预测值做为额外的输入值，组成新的历史数据，则，依次类推，直至预测出需要长度的数据值。

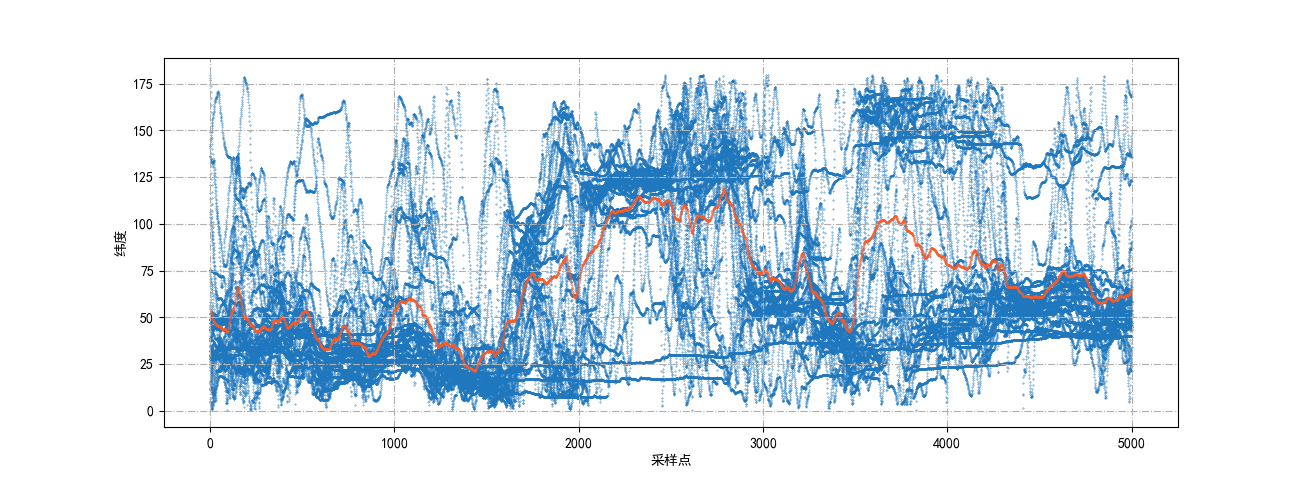
可以看出，随着预测时间的增加，预测误差将会逐渐增大，为了进一步提高预测准确率，我们参考基于内容显著性检测的视点预测方法，这种方法根据视频内容的显著性来辅助预测，由于视频各个区域内容的显著性不同，人们在某个时刻观看某个区域的概率也不相同，当该区域视频内容较为显著性时，用户观看的可能性也较大，所以，多个用户的观看轨迹可以从侧面反映视频内容的显著性。为了验证这一假设，我们从数据集中随机抽取了2个视频文件，绘制了48名用户观看这两个视频的轨迹散点图，如图 所示，其中蓝色散点为用户的观看位置，红色散点为平均值。





(a) 观看视频1的轨迹图





（b）观看视频2的轨迹图

图 48名用户观看轨迹图

从图（a）中可以看出，视频最开始时，由于用户

用户因此，我们结合观看同一视频的其他用户的观看轨迹，对LSTM模型初步预测出来的数据进行矫正，提高预测准确度。

根据上述分析，我们认为无论是有一个还是多个感兴趣区域的全景视频，其他用户的经纬度数据对当前用户的经纬度数据都有着一定的参考价值。因此，我们提出一种结合其他用户视口数据的预测算法。

为了衡量其他用户历史数据的参考重要性，我们引入基于tile的传输机制，将视频划分为16\*8块tile，记第i行j列的tile的观看次数为，初始值设为0。首先，我们根据当前用户的历史数据分别预测出接下来的经纬度值，确定视口区域以及视口所覆盖的tile块，这些区域每块tile的观看次数加，然后根据其他用户在同一时刻的视口区域位置，将覆盖的tile的观看次数依次加，最后，每个tile的观看次数即为两者之和，计算公式如下：

其中，取值0或1，表示根据LSTM模型预测出来的区域是否覆盖该tile，n为视口区域覆盖该tile的其他用户的个数。和取值的比例关系代表了其他用户数据的参考性，例如，即我们认为最终预测的视口区域由LSTM模型预测出来的视口区域和其他用户的视口区域以相同的重要性共同确定。考虑到LSTM预测出来的视口数据的准确度会随着预测时间的增加而降低，的取值也应该随着的变化而变化，为了减小后续计算的复杂度，我们定义和的计算公式如下：

其中，N为其他用户的总个数，为系数，可以根据实验结果对的取值进行调整以提高预测准确度，为向下取整函数。



### 图 tile观看次数示意图

图 为一帧视频的观看次数示意图，假设视频帧划分为16\*8块tile，则视口区域占6\*4块tile，。计算出所有tile的观看次数后，我们可以获得一个二维数组，然后计算出6\*4块tile的和最大区域即为最终预测的视口区域所占tile，如图中红色区域所示。

2.3 实验与结果分析

本节介绍视口预测实验的结果和分析，首先介绍衡量预测算法准确度的评价指标，然后对比基于LSTM模型的预测算法与其他基准算法在不同评价指标下的预测准确度。

2.3.1 实验过程

2.3.1 评价指标

为了对比我们提出的算法与其他基准算法的预测准确度，本节使用两种评价指标，经纬度绝对差值和tile重叠比。

（1）经纬度绝对差值

为了衡量LSTM模型的预测准确度，我们使用同一时刻真实的经纬度与预测的经纬度的绝对差值来代表预测误差，由于全景视频的特性，到在经度上，-150度和150度直接取差值为300度，但实际两者只差了60度，因此，我们使用文献[]中定义的角度差值的计算方式，公式如下：

其中，代表角度和的差值，为取模运算，基于此，我们定义预测时刻经纬度的绝对差值如下：

（2）tile重叠比

由于我们提出的预测算法最终预测的是视口区域所占tile的情况，所以我们使用同一时刻预测的视口所占tile在真实视口所在tile中的重叠比做为最终的预测准确度指标。

2.3.2 实验与结果分析

为了全面评估基于LSTM模型的预测算法的预测准确度，我们对两个实验参数进行设定：（1）历史时间窗口，设置三个参数[0.5s，1s，1.5s]；（2）预测时间窗口，为了验证LSTM模型在长期预测时的性能优越性，设置六个参数[1s，2s，3s，4s，5s，6s]。我们分别将经纬度时间序列按照为一组进行划分，采用移动窗口思想最终产生若干组子序列，为了消除不同类型视频对预测结果的影响，我们采取交叉验证的方式。

为了对比预测准确度，我们研究了两种回归方法，线性回归-LR和非线性回归-MLP，将时间和经纬度值视为独立变量，目标是根据历史窗口的数据拟合出回归函数，然后根据预测窗口得出预测数据，我们对经纬度分别建模，对于MLP，我们采用了一个简单的结构，隐藏层含有3个神经元，并使用relu做为激活函数，ADME做为优化函数。最后，我们使用移动平均算法做为基准算法，即使用历史窗口中的数据的平均值做为整个预测窗口中的预测值。

2.4 本章小结

全景视频的视口预测是全景视频传输优化问题的重要前提，优秀的预测算法可以显著降低全景视频的传输数据量。本节针对现阶段视口预测算法长时间预测准确度明显下降的问题，提出了一种基于历史观看轨迹的预测算法。首先，该算法根据当前用户的历史观看轨迹，使用LSTM模型对轨迹的经纬度分别预测，得到初步预测结构。为了进一步提高预测准确度，在视频服务器存储了观看同一视频的其他用户的轨迹数据的情况下，我们计算其他用户在全景帧每块tile的观看次数，再结合LSTM模型的输出结果，综合计算出预测窗口所覆盖的tile块。实验结果表明，基于当前用户和其他用户历史观看轨迹的预测算法比起其他常用的预测算法有着较高的准确度。

# 第四章 总结与展望

4.1 工作总结

随着计算机科学和多媒体技术的发展，全景视频作为一种全新的信息媒介以其沉浸式构想性的特点给人们带来了身临其境的观感，并逐渐应用到各个领域中。然而，全景视频数据量大、对时延要求严格的特性也给现有的传输网络带来了一些挑战，如何优化全景视频的传输一直是受人关注的研究领域。因此，本文从降低传输数据量和高效调度无线资源两个方面进行改进，主要工作研究与成果如下：

（1）为了降低全景视频的传输量，本文首先研究了视口预测算法。考虑到基于内容分析的预测算法的复杂性，本文只聚焦于基于历史轨迹的预测算法，该场景可以归于时间序列预测问题，在介绍了一些常见的时间序列预测算法之后，基于视口的二维特性，即经纬度，初步提出了基于序列到序列模型的预测算法，在数据集上初步验证了该算法的预测准确度高于其他一些算法，然后，为了进一步提高预测准确度，本文分析了观看同一个视频的其他用户的视口轨迹，发现轨迹之间的相似度极高，基于此前提，本文结合其他用户的视口数据对初步预测的结果进行矫正，最终，实验结果证明联合其他用户的预测算法在长时间预测时可以有效提高预测准确度。

（2）减少全景视频的传输量之后，本文又考虑了LTE网络中基站侧的无线资源调度。首先介绍了LTE的相关理论，包括网络架构和协议栈架构，详细阐述了LTE系统中的关键技术。然后介绍了无线资源块的概念，给出了资源调度和分配流程。针对全景视频这种对时延极为敏感的业务，本文研究了经典的调度算法的优缺点后，基于强化学习中的Q-Learning算法，提出了一种创新的下线资源调度算法DSQL，该算法根据动态变化的网络状态和用户状态，使用Q-Learning算法在每个传输间隔选择不同的调度策略，最终达到系统性能最大化目的。为了验证DSQL算法的性能优越性，本文介绍了仿真平台LTE-Sim的相关功能和仿真流程，通过仿真实验，从系统吞吐量、系统丢包率、时延和用户公平性几个方面综合对比了DSQL和经典算法的性能，最终验证了DSQL算法的优越性，尤其在时延敏感的实时业务调度方面。

4.2 未来展望

全景视频传输优化是一个庞大的研究领域，设计到计算机视觉、数据挖掘、无线网络传输优化等多个研究学科，本文只在现有研究基础上做了部分改进，仍有一些问题需要未来做进一步深入研究：

（1）本文只聚焦于基于历史轨迹的视口预测算法，没有充分考虑到影响视口的多个因素，如视频内容的多样性和用户的观看兴趣与习惯，没有建立充分完善的视口模型；在对视口预测时，分别预测了视口中心位置经纬度的规律，没有进一步探索其余预测模式，如基于区块的预测，这些因素都将会影响预测准确度，后续需要综合考虑深入研究。

（2）在资源调度方面，本文只是使用数据量较大这一个特征来代表全景视频业务进行调度，没有考虑全景视频的其他特性，如视口区域数据的重要性明显高于其他区域的数据，应当提高这些数据的优先级。同时在仿真过程中只考虑了单小区多用户的场景，而实际中更多的是多小区场景，存在小区间干扰等现象，之后应当更加紧密结合全景视频的特性，针对多小区的场景，进一步优化资源调度算法的性能。

（3）本文在研究资源调度算法时聚焦于LTE网络，而在研究后期，5G网络的发展逐渐加快，相关技术研究也不断完善，未来5G网络的覆盖将改善全景视频的传输问题，因此后续可以考虑结合新型网络架构与技术，进一步优化全景视频的传输。

# 致谢

时光如白驹过隙，转眼间充实又紧张的三年研究生生涯即将告一段落，之后将要告别校园生活步入职场社会，论文行文至此，内心不禁思绪万千。回顾这三年来，收到过许多人的关心与帮助，在此向激励我、支持我、帮助我，让我不断成长的老师、家人、同学和朋友致以我最真诚的感谢。

首先由衷地感谢我的导师刘雨老师。刘老师在学术上对我帮助甚多，从如何高效地查找科研资料、如何正确地阅读参考论文到如何进行科研实验，指导我从只会学习的本科阶段顺利过渡到可以独立科研的研究生阶段；针对我的研究课题，刘老师总能提出关键性问题，在研究瓶颈期引导我开拓思路，正是刘老师的谆谆教导才让我取得了一定的科研成果；此外，刘老师严肃认真的科研态度也不断影响着我，真的很幸运可以遇到这样一个良师。

其次，我要感谢望育梅老师。在科研学习与研究生生活中望老师也给予了我许多的支持与帮助。在准备发表论文成果期间，望老师细致耐心地辅导我撰写论文，纠正论文中的语法错误，给我提出了许多有用的修改建议。另外我也要感谢张琳老师，张老师积极组织我们参加科研讲座、参与科研项目、阅读最新研究成果，正是因为有张老师的督促和鼓舞才扩宽了我科研的方向。当然，先进网络技术实验室的每一位老师都给予过我各种各样的帮助，再次表达我最诚挚的感谢。

由衷感谢实验室的每一位小伙伴。感谢李东海和吕扬师兄，两位师兄的热情与风趣让我迅速融入了1007这个大家庭，感谢你们在科研、工作和生活上给与我的建议，祝福你们工作顺利；感谢小伙伴刘晓钰、徐安越、范文晋、于越和胡煜翔，是你们的陪伴才让紧张枯燥的科研生活变得精彩缤纷，尤其感谢刘晓钰同学，我们工位紧挨一起，一起科研，一起吃饭，一起出差开会，我们一起度过了许许多多快乐的时光，衷心祝福你们可以心想事成；感谢可爱的师弟师妹们，你们的加入让我们这个家庭不断扩大，也十分感谢你们对师姐的帮助，也祝福你们在科研道路上越走越顺畅，产出许多优秀的科研成果。

感谢我的室友宋越、高雅、王丹頔、苏明兰和张思洋，我们来自五湖四海，十分幸运与你们成为了亲密无间的室友，感谢你们在我科研困顿的时候，在我找工作迷茫的时候给予我最坚定温暖的安慰与支持，感谢你们的参与给我的研究生生涯留下了无数快乐的美好回忆，祝福你们未来可期，前程似锦。

最后，要感谢默默支持我关心我的家人，让我可以毫无顾忌的追求自己的梦想，祝福家人身体健康，万事如意。