UHD World Association
世界超高清视频产业联盟

体育场景数实融合技术研究报告

Digital-physical Fusion Based Video Technology for Sports

UHD World Association
世界超高清视频产业联盟

UHD Word Association
www.theuwa.com

前 言

本文件由UWA联盟智慧体育专题组制订，并负责解释。

本文件发布日期：2025年06月11日。

本文件由世界超高清视频产业联盟提出并归口。

本文件主要起草单位：

中国移动通信集团有限公司、咪咕文化科技有限公司、北京大学、广东广播电视台、北京百度网讯科技有限公司、马栏山音视频实验室、鹏城实验室、杭州当虹科技股份有限公司、华为技术有限公司、上海海思技术有限公司、北京大学深圳研究生院、上海交通大学、四开花园网络科技(广州)有限公司、深圳市沉浸视觉科技有限公司、广东图盛超高清创新中心有限公司、赛因芯微（北京）电子科技有限公司、深圳市奥拓电子股份有限公司

本文件主要起草人：

贝悦、李琳、马思伟、王荣刚、王苫社、陈望都、王琦、徐嵩、高峰、贾川民、陈丽丽、丁凌、单华琦、毕蕾、尹茜、刘烜奕、于路、罗映辉、刘川、马学睿、高铁铸、赖亚军、徐异凌、梁小明、成六祥、胡颖、李维、芦超、杜江、王勇、张兴民、陈博、张玮 、陈家兴、梁超翔、梁锋、吴英、潘兴浩、谢于贵、李康敬、许海滨、郭佩佩、宋泽田、宁泓博、唐浩程、尹馨慧、孙睿涵

免责说明：

1， 本文件免费使用，仅供参考，不对使用本文件的产品负责。

2， 本文件刷新后上传联盟官网，不另行通知。

目录

1. 引言

随着数字技术的飞速发展，数实融合技术凭借其独特优势在多个应用领域日益受到广泛关注。**该技术通过将真实拍摄的影像内容与虚拟现实和增强现实等生成的三维模型有机结合，旨在为用户提供更加真实、具备多自由度交互的沉浸式观赛视觉体验。**在现有的相关技术中，XR虚实制作技术、三维沉浸式技术以及数实融合技术常被用作比较对象，它们在实现路径和应用场景上既有交集，也各具特色。如表1-1所示，XR虚实制作技术涵盖虚拟现实（VR）、增强现实（AR）和混合现实（MR），强调用户的沉浸感和交互体验，通常需要借助特定的硬件设备，如VR头显或AR眼镜来实现。三维沉浸式技术则侧重三维立体感和全景视角的呈现，主要依赖高分辨率相机阵列和全景相机等设备，力求营造具备空间深度的沉浸式视觉效果。相比之下，**数实融合技术的独特之处在于，它能够通过普通屏幕或高性能显示设备（如AR/VR头显）将数字内容与现实场景自然结合，既支持无特殊设备的便捷体验，也可通过专业设备提升交互性和沉浸感。**

表1-1 XR虚实制作技术、三维沉浸技术及数实融合技术特点

类别	特点	硬件设备	应用场景	优势	不足
XR虚实制作技术	支持虚实互动 沉浸感强烈	依赖专业设备 如VR头显和AR眼镜	虚拟会议、 虚实游戏等	高互动性，支持复杂虚拟场景	设备门槛高 普及性差
三维沉浸技术	强调三维立体感和全景视角	依赖高分辨率相机阵列、全景相机、3D显示设备	旅游、教育等	视觉体验丰富 适合三维场景呈现	交互性较弱
数实融合技术	结合虚拟内容和真实场景	支持普通屏幕，也支持高分辨率摄像头、头显等专业设备	体育转播、 在线教育等	设备依赖性弱 便捷性较好	技术复杂性高 实时处理要求高

随着数字化技术的不断突破，三维建模与虚拟内容生成技术在影视、游戏、元宇宙、在线教育以及体育赛事转播等领域的应用日益广泛。这些领域不仅对高保真度的视觉效果提出了严格要求，还对实时交互能力有了更高的期待，以满足用户不断升级的体验需求。数实融合技术与人工智能、虚拟现实、超高清视频等产业发展深度关联，为这些领域提供了强有力的技术支持。目前我国数实融合技术相关的政策框架正在逐步建立，主要是促进和鼓励相关技术的培育和发展，以及拓展在多元场景下的应用。2021年3月，国家将虚拟数字技术的发展纳入“十四五”规划，明确了虚拟现实技术创新对实现产业升级和建设技术强国的重要意义。从产业层面来看，数实融合技术能够将虚拟内容与现实场景有机结合，推动影视、游戏、元宇宙和体育赛事转播等行业的数字化转型和升级，促进产业链协同发展。从技术层面来看，该技术涉及虚拟内容生成、三维建模、视频处理与

编码以及数据传输等核心领域，并将其与生成式人工智能和计算机图形学进一步融合，以应对现阶段技术挑战，并提出标准化发展建议。

然而，现有的视频技术白皮书等多聚焦于单一模态内容，如超高清视频和沉浸式技术等，已难以适应不断创新发展的生成式人工智能和计算机图形学与多模态内容结合带来的新需求。此外，不同三维模型内容和视频格式对技术处理方法的多样性要求进一步凸显，现有研究缺乏通用性。为此，**本技术研究报告旨在为体育场景中数实融合相关的产业发展和能力建设提供全面指导，促使行业在数实融合视觉计算技术创新方面取得进展，为技术研究、内容制作和系统开发提供全面支持。**最终为引领行业发展、推动技术创新和提高用户体验做出重要贡献。

2. 数实融合技术体系概述

数实融合技术体系旨在将数字世界与物理世界相结合，通过将多种数字技术生成的内容与真实场景进行深度融合，为用户提供沉浸式体验。如图2-1所示，数实融合技术体系从关键技术、系统解决方案、业界实践等维度构建完整生态，促进虚拟世界与真实场景的紧密结合，推动数实融合相关的产业发展。

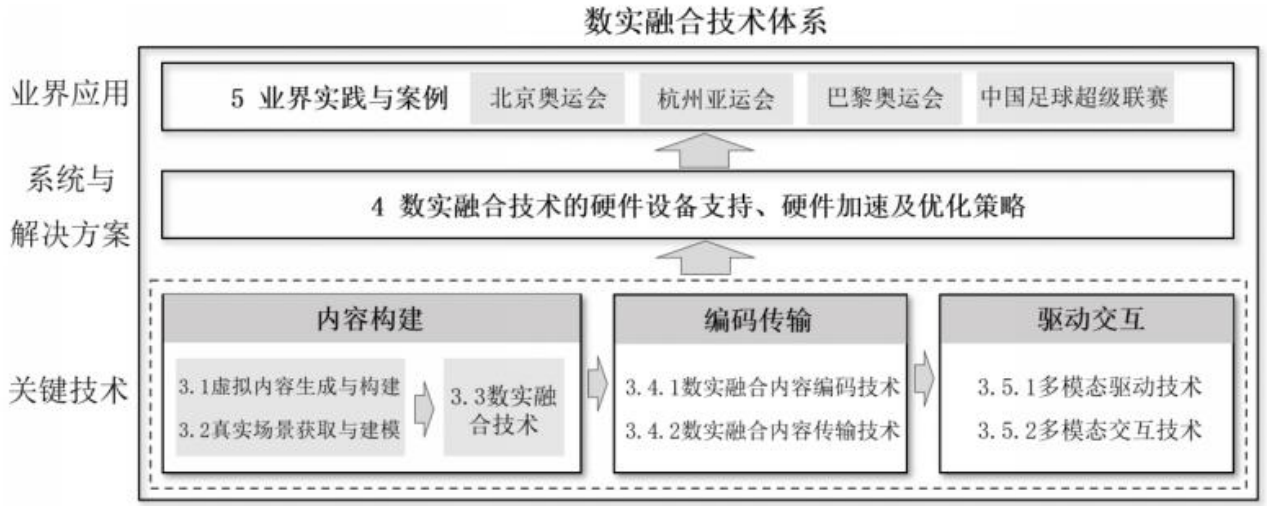

图2-1 数实融合技术体系框图

首先，**关键技术是数实融合体系的核心支柱**，可分为三大部分。第一部分包括虚拟内容生成与构建、真实场景获取与建模以及数实融合技术。虚拟内容生成依托3D建模、虚拟环境构建和实时渲染工具，为数字世界搭建虚拟空间；真实场景获取与建模利用高精度传感器和智能数据处理技术，完成现实世界的数字化表达；数实

融合技术则通过数据流整合和多模态呈现，实现虚拟与现实的无缝衔接。第二部分是数实内容的编码传输技术。该部分聚焦于压缩编码、低延迟传输协议和分布式处理方法，以保障多终端设备和复杂网络环境中的内容传递效率，提供流畅的用户体验。第三部分是多模态驱动与交互技术，通过语音、手势、眼动追踪等多种交互方式，结合AR/VR设备，构建自然直观的人机交互界面，进一步提升数实内容的应用深度与用户体验。

其次，**系统与解决方案为数实融合技术的实现提供了全面支撑。**硬件支持是该体系的重要基础，包括高性能GPU、3D显示设备、深度摄像头等，用于数实内容的生成、建模与实时渲染。此外，优化策略如边缘计算和分布式渲染显著提升了实时处理能力和资源利用效率，同时降低了系统功耗。结合硬件加速和软硬件协同优化，这些方案为数实融合技术在复杂应用场景中的落地提供了技术保障。

最后，**数实融合技术的实际应用已覆盖多个行业领域，并展现了广阔的发展前景。**在北京冬奥会、杭州亚运会及巴黎奥运会上，数实融合技术为赛事直播、场馆虚实互动和沉浸式观赛体验提供了技术支持，成功展示了其商业价值与社会效益。

未来，随着人工智能、5G/6G通信技术和边缘计算的快速发展，数实融合技术将在实时渲染、沉浸式交互和跨平台协作等领域取得更大突破。与此同时，国家层面正加速推进政策与标准化建设，通过技术规范的统一与生态体系的完善，助力数实融合技术的规模化应用与全球化推广。

通过构建完善的数实融合技术体系，虚拟与现实的界限将被不断打破，技术与行业需求的结合将更加紧密。这为沉浸式应用创新提供了技术支撑，也为未来数实融合技术在体育场景中的发展开辟了更广阔的空间。

3. 数实融合关键技术

3.1 虚拟内容生成与构建

虚拟内容生成与构建技术是数实融合领域中的核心技术之一，通过数字化手段创造逼真且互动性强的虚拟世界。随着计算机图形学、深度学习和生成式人工智能等技术的发展，虚拟内容的生成效率和质量得到了显著提升，广泛应用于虚拟现实、影视制作、游戏开发、以及数字孪生等领域。

3.1.1 图像与视频生成技术

随着深度学习技术的发展，基于数据驱动的深度生成网络已经成为图像和视频生成的主流方法。这类技术

的核心思想是：通过让神经网络从大量真实图像和视频中学习，掌握它们的视觉特征和规律。之后，网络就可以根据这些学习到的“模式”来生成新的图像或视频内容。如果再加入用户的控制指令，比如想生成什么风格、什么内容的画面，系统还能根据这些要求有针对性地生成高质量的结果。该技术已经广泛应用于虚拟内容创作、视频特效和多媒体应用中。本文主要关注文本图像视频生成技术，即在给定文本描述的条件下，生成与文本内容一致的图像或视频，实现文本到图像或视频的高质量一致性生成。

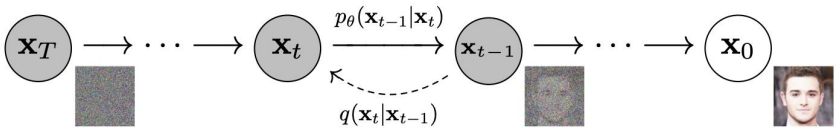

图3-1 图像扩散模型前向（加噪）与反向（去噪）过程[3]

为了实现视觉上真实且语义准确的图像生成，研究者提出了多种模型和方法，其中主要的深度学习方法有生成对抗网络[1]、自回归模型[2]和扩散模型[3]，其中，扩散模型是如今最主流和强大的生成模型。能通过训练模型学习文本和图像之间复杂的映射关系，实现逼真的图像生成，并且能够灵活地产生数据集中未出现过的视觉内容，推动了该领域的快速发展[4]。

扩散模型是一种强大的生成模型，它通过模拟物理扩散过程来生成新的数据样本。这种生成模型的核心包括两个主要步骤：前向扩散过程和反向扩散过程。在前向扩散过程中（见图3-1中的实线箭头），系统会一步步地往原始图像中添加噪声，直到图像完全变成类似“雪花点”一样的纯噪声图。这一过程可以看作是一个有顺序的过程，每一步都在前一步的基础上继续加噪声，直到看不出原图为止。而在反向扩散过程中（见图3-2中的虚线箭头），模型的任务就是反过来操作——从这张全是噪声的图像开始，一步步地“去噪”，最终还原出原来的图像，或者生成一个全新的、清晰的图像。这个过程其实就是模型在“猜测”原图长什么样，并试图重建出来。为了让模型学会如何加噪声和去噪声，它需要通过大量数据进行训练。这个训练过程用到了深度学习中的反向传播算法和变分推断技术。经过反复学习，模型最终可以掌握图像中蕴含的规律，从而具备生成新图像的能力。

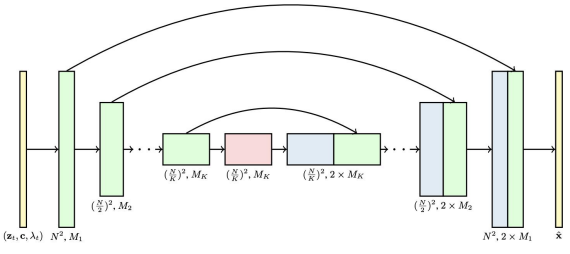

图3-2 视频扩散模型的3D-U型网络结构 [5]

视频内容生成同样也发展出了生成对抗网络[1]，自回归模型[2]等深度生成网络，但一直到2022年以来，视频扩散模型[5]提出，才革命性地推动了视频内容生成方法的发展。该方法采用3D-U型网络对视频中的时空噪声进行建模，将扩散模型成功应用于视频内容生成。这种方法不仅是在原有图像扩散模型的基础上进行改进，还进一步探索了如何同时处理图像和视频数据的训练方式。它的一个重要优点是：能减少训练中小批量数据带来的不稳定性（也就是梯度的波动），从而让模型训练得更快、更稳定。为了让模型能生成更长时间、分辨率更高的视频内容，这种方法还引入了条件采样技术。该技术可以帮助模型在空间（画面细节）和时间（帧数连续性）两个方面扩展生成效果。整体来看，这种改进方法不但能生成更高质量的视频，还能保证视频在时间和空间上的连贯性，比以前的方法有明显进步，也为今后的视频生成技术打下了坚实的基础。

3.1.2 三维建模与渲染技术

三维建模与渲染技术是虚拟内容构建中的关键环节，旨在创建和展示高度真实的三维场景、物体和人物模型。随着技术的进步，自动化建模和智能渲染的结合，使得三维虚拟世界的构建变得更加高效且真实，能够自动化生成复杂三维模型和虚拟环境，同时利用实时渲染技术（如光线追踪、全局光照）优化三维内容的显示效果和表现力，主要分为三维场景建模与对象构建，纹理映射与材质设计以及实时渲染等技术[6]。

（1）三维场景建模与对象构建：三维建模技术使用多边形网格建模、曲面建模等方法构建虚拟物体和环境。近年来，深度学习和AIGC技术的结合使得复杂的三维模型生成变得更加自动化。例如，用户也可以通过简单的文本描述来创建复杂的三维模型。这种方法通常依赖于深度学习、自然语言处理（NLP）和计算机视觉技术的结合[7]。文本驱动的三维物体生成方法应用领域广泛，包括游戏设计、虚拟现实、建筑可视化和教育工具等。

（2）纹理映射与材质设计：纹理映射是三维动画设计中实时渲染技术的一种重要方法，可以将二维图像或纹理映射到三维模型的表面，增加模型的细节和真实感。在实时渲染中，纹理映射的优化可以提高渲染速度和图像质量，使呈现效果更加逼真和流畅[7]。结合AIGC技术，可以自动化生成高度逼真的纹理与材质。例如，AI可以根据光照、表面反射等物理属性自动调整材质，提升虚拟物体的细节和视觉效果。

（3）实时渲染技术：光线追踪是三维动画设计中实时渲染技术的一种重要方法，模拟光线在场景中的传播和相互作用，以生成逼真的图像。随着光线追踪（Ray Tracing）和全局光照（Global Illumination）技术的成熟，虚拟内容的视觉表现力得到了显著提升[8]。结合AIGC技术，渲染引擎可以在保留高质量图像的同时，

动态调整渲染效果，提升渲染速度和资源使用效率，尤其在实时应用中发挥重要作用。

通过图像、视频生成技术与三维建模和渲染技术的紧密结合，虚拟内容生成与构建技术正推动数字世界的发展，带来更高效、更真实的虚拟体验，并为未来的数实融合应用提供技术保障。

3.2 真实场景获取与建模

真实场景获取与建模技术是实现物理世界数字化的关键环节，为沉浸式应用提供了真实体验感的基础。通过先进的传感设备和算法技术，能够高效地采集场景中的几何、纹理和光照信息，并进行精准的数字化建模。该技术涵盖从数据采集到三维模型生成的完整流程，广泛应用于数字孪生、虚拟现实（VR）、增强现实（AR）以及影视制作等领域。真实场景采集技术侧重于高质量数据的获取，而三维重建技术则致力于将采集的数据转化为逼真、可交互的三维模型，两者相辅相成，共同推动数实融合技术的创新发展。

3.2.1 真实场景采集技术

除了上述虚拟内容外，真实场景数据也是数实融合内容技术的核心之一。通过对真实场景的完整捕捉和建模，让用户获得一种身临其境的视觉体验。作为数实融合内容制作的第一步，数据采集决定了后续重建、渲染以及最终显示的质量和效果。近年来，随着计算机视觉、深度学习等技术的发展，三维内容的采集技术也不断演进，从传统的双目相机采集逐渐扩展到多视点阵列、全景相机和深度相机等多种方式，每一种技术都有其独特的应用场景和优势。

图 3-3 RGB-D 相机、双目相机以及三维激光扫描仪

（1）双目相机采集是一种相对成熟且广泛应用的三维数据获取方式，它通过模拟人类的双眼视觉，以两台摄像机同时拍摄略有偏移的同一场景，从而获取丰富的视觉信息。这种采集方式的核心在于视差的计算，通过对两个摄像头拍摄的图像进行比对，系统能够精确计算出场景中物体的深度，从而形成具有立体感的图像。

（2）多视点阵列通过将多个摄像头按照特定的几何方式排列，从多个视角同时采集同一场景，从而为三维重建提供更丰富的数据支持。阵列相机不仅扩展了视角的数量，而且提高了场景的覆盖范围，从而更好地还原场景的三维结构。

（3）全景相机则进一步提升了观众的沉浸感。全景相机通常采用多个摄像头或鱼眼镜头的组合布局，通

过多视点拼接技术实现360度无死角的场景采集。这种方式能够为观众提供全方位的观看体验，使他们在观看时可以任意选择视角，仿佛置身于真实场景之中。全景相机常用于虚拟现实和自由视点视频应用，通过结合立体投影模型（如全方向立体投影），为左右眼分别生成全景图像，从而实现完整的立体观看效果。

（4）深度相机作为三维信息采集的关键设备，也在数实融合视频中发挥着重要作用。深度相机（如RGB-D相机）结合了传统彩色摄像头和深度传感器，可以同时捕获场景的颜色和深度信息。深度相机的工作原理既可以基于被动方法，如双目立体视觉，也可以基于主动方法，如结构光或TOF（飞行时间）相机。这些技术能够实时提供每个像素的深度信息，为后续的三维重建和渲染提供了精确的几何数据支持。

除了这些基于视觉的采集技术，激光扫描作为一种高精度的三维采集方式，通过发射激光并记录反射时间来精确计算物体的三维坐标，从而生成密集的点云数据。这些点云数据经过后期处理，可以生成非常精确的三维模型，为场景的重建和细节复刻提供了重要支持。关于常见的真实场景采集技术的工作原理、优缺点和应用场景见表3-1。

表3-1 真实场景采集技术

采集技术	工作原理	优点	局限性	典型应用
双目相机	模拟人类双眼视觉，用两台摄像机拍摄略有偏移的同一场景，通过视差计算获取物体深度信息。	成本低廉，硬件易部署，适合中小型和静态场景的内容采集。	复杂场景中视差算法难以恢复细节，部分立体信息可能缺失。	中小型场景采集、静态场景建模
多视点阵列相机	多摄像头按照特定几何排列，从多个视角采集同一场景，为三维重建提供丰富数据支持。	扩展视角数量，提高场景覆盖范围，适合大场景和动态场景捕捉。	硬件成本高，数据处理复杂，需高效融合算法支持。	大场景动态捕捉、自由视点视频直播
全景相机	采用多个摄像头或鱼眼镜头组合，通过拼接技术实现360度场景采集，为左右眼生成全景图像实现立体观看效果。	提供全方位沉浸式观看体验，适合虚拟现实和自由视点视频应用。	拼接存在缝隙和失真，动态场景对同步与图像融合要求高。	虚拟现实内容采集、全景视频制作
深度相机	结合彩色摄像头与深度传感器，基于双目立体视觉、结构光或TOF（飞行时间）等技术采集颜色和深度信息，提供精确的几何数据支持。	实时获取深度信息，适合动态场景和实时反馈应用。	受环境光和测量距离限制，分辨率较低。	动态场景建模、实时交互场景
激光扫描	通过发射激光并记录反射时间计算三维坐标，生成密集点云数据，进一步生成精确的三维模型。	高精度，适合高细节场景的采集，如建筑测量和文物保护。	成本高，对环境依赖性大，无法处理透明或高反射表面。	建筑测量、工程设计、文物保护

3.2.2 三维重建技术

三维重建技术是数实融合体系中至关重要的环节，它通过对采集到的数据进行处理和分析，将平面的图像

信息转化为具有立体感的三维模型，为观众提供高度沉浸的观看体验。当前，三维重建的发展主要沿着三种技术路线展开：基于传统几何方法的多视图三维重建（如SFM和MVS）、基于点云与网格表示的几何建模方法，以及近年来兴起的基于神经表示的重建方法（如NeRF和3D Gaussian Splatting）。这三种技术路线在重建思路、数据表示、渲染效率与最终应用上各有优势，相互补充，共同推动了三维视频从平面图像向空间立体结构的跃升。

首先，基于传统几何的方法在三维重建中依然发挥着重要作用。这类方法以传统计算机视觉算法为核心，依赖多视角图像进行三维几何信息的还原。重建过程通常由运动恢复结构（Structure from Motion, SFM）和多视图立体视觉（Multi-View Stereo, MVS）组成，前者用于提取稀疏三维结构，后者实现稠密建模。

SFM基于多个视角图像，从多张不同视角的图像中提取特征点并进行匹配，通过几何校验剔除错误匹配项，并通过光束平差（Bundle Adjustment）优化相机姿态和三维点位置的精度。稀疏点云虽无法直接用于渲染，但为后续稠密重建奠定了基础。

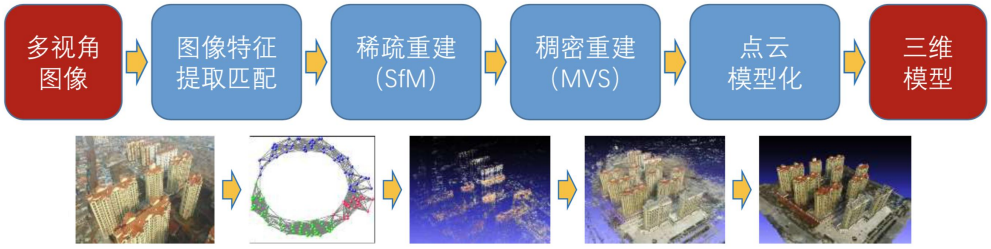

图3-4 基于图像的重建算法流程

多视图立体视觉（MVS）[10]是实现稠密重建的重要技术手段之一。MVS通过多台相机从不同角度拍摄的图像，计算场景中每一个点的深度，从而生成更为精细和完整的三维模型。MVS的方法可以分为体素重建、点云扩散以及深度图融合等多种技术路线：1）体素重建法通过对整个三维空间进行划分，将每个小体积单元（体素）进行计算，以确定其是否属于场景中的物体。这种方法具有很高的精度，但对计算资源的要求非常高，尤其是在需要高分辨率时，体素数量呈指数增长，导致计算复杂度急剧增加。2）点云扩散法则基于稀疏点云，通过扩展每个点在不同视角下的位置来生成稠密点云，其优势在于点的分布较为均匀，适合大场景的重建。3）深度图融合是另一种实现稠密重建的重要方法。它通过估计每个视角下的深度图，并将所有深度图融合到一个三维点云中，以生成完整的三维模型。这种方法利用GPU的并行计算能力，可以在视角数量较多的情况下高效完成重建任务。相比其他方法，深度图融合生成的点云密度较高，且在细节恢复方面表现优秀，是目前应用较

为广泛的稠密重建技术之一。此外，深度学习技术的应用也为MVS提供了新的思路，特别是在特征匹配和深度估计方面，通过卷积神经网络等模型的引入，显著提升了重建精度和鲁棒性。

第二，基于点云与网格表示的几何建模方法。由于重建后的结果往往以点云形式呈现，因此点云与网格结构的构建成为进一步提升视觉质量的重要步骤。点云是三维空间中离散点的集合，常由激光雷达或深度相机获取。为了得到完整的三维模型，需对点云进行配准、滤波与拟合，提升其空间一致性与准确性。为了更适合可视化与渲染，点云通常会被转化为三角网格结构。网格重建通过在点云中连接相邻点形成面片，实现从离散点集到连续曲面的转化，使得模型细节更加完整，表现更具真实感。在重建过程中，基于图像和基于几何的两类重建方法各具优势。基于图像的方法，例如光场渲染，主要通过采集大量视角下的图像，并通过图像融合来实现三维场景的生成，其优点在于不需要对场景的几何信息进行显式建模，能够在一定程度上避免几何重建中的伪影和不连续问题。而基于几何的方法则直接通过计算三维结构来生成场景模型，如点云和体素表示等，适合需要高度精确的场景重建。近年来，混合表示方法逐渐流行，这种方法结合了图像和几何信息的优势，利用多视点与深度图的结合，不仅减少了对视点数量的依赖，同时也提高了生成视点的灵活性和视觉效果。

另外，神经辐射场（Neural Radiance Fields，NeRF）[11]是一种基于神经网络的三维场景表示方法（如图3-5所示）NeRF 通过神经网络端到端地学习每个空间点的颜色和密度，能够在没有显式几何建模的情况下完成高质量三维重建。其输入为三维坐标和观察方向，输出为对应点的辐射强度与颜色值，能够逼真地呈现光照、阴影、反射等复杂视觉效果。

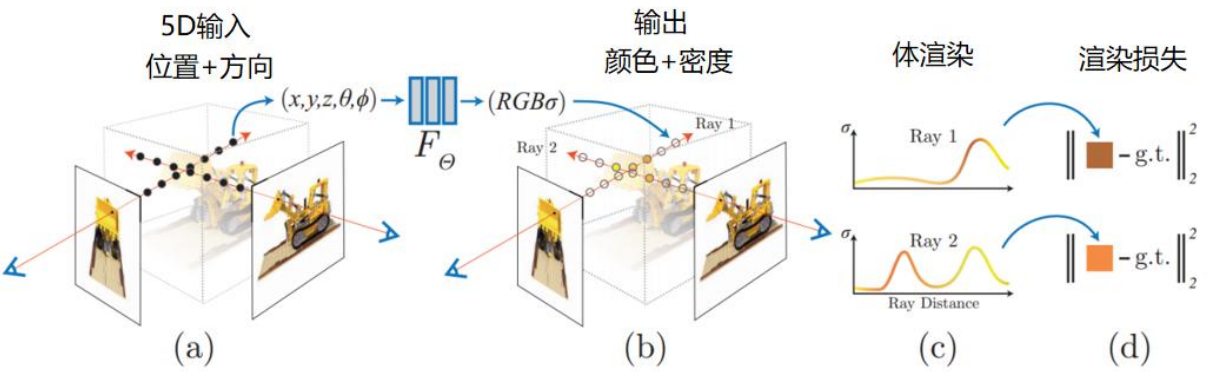

图 3-5 NeRF 三维重建框架[11]

3D高斯（3D Gaussian Splatting，3DGS）[12]则提供了另一种三维重建方式（如图3-6所示），进一步优化了渲染效率。3DGS 将场景表示为一组高斯分布点，利用可微光栅化实现快速投影与渲染，避免了传统 MLP 渲染速度慢的问题，尤其适合应用于数字人建模和动态场景重建。

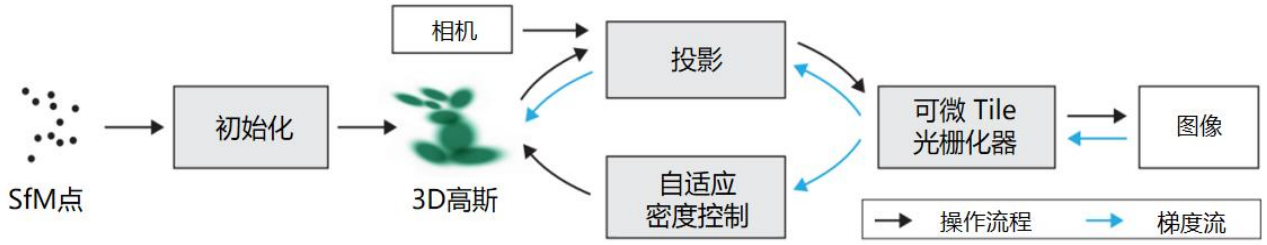

图 3-6 3D Gaussian Splatting 重建框架[12]

三维重建技术正朝着“几何建模+神经表达+结构优化”融合发展的方向演进。无论是基于图像的几何方法、基于结构的点云网格方法，还是基于学习的神经场方法，都在不同应用场景中发挥着独特价值，推动数实融合的三维内容生成迈向更高的质量和效率。

3.3 数实融合技术

数实融合技术的本质在于通过虚拟生成内容与真实场景数据的深度整合，构建出具有空间感知、环境响应与交互能力的融合内容，从而满足包括沉浸式观赛、智能制造、数字孪生等在内的多种新兴场景需求。本节将围绕数实融合的关键实现路径展开，具体包括虚实结合技术和背景替换技术两大类。前者关注如何在真实场景中自然嵌入虚拟内容，实现几何、光照、材质等方面的一致性；后者则着重于前景提取与虚拟环境合成，提升虚拟场景的可塑性与沉浸感。

3.3.1 虚实结合技术

虚实结合技术作为数实融合的核心，旨在确保虚拟元素（如3D模型、虚拟角色、动画等）与真实环境之间的无缝整合。为了实现这一点，计算机图形学技术发挥了至关重要的作用，尤其是在光照、阴影、反射等方面的处理。这可以通过光照模型和反射模型实现，确保虚拟元素在真实环境中看起来自然。相关技术如下所示：

（1）光照一致性：真实世界中的光照变化会影响物体的外观，因此虚拟物体的光照也必须根据真实场景中的光照条件进行调节。通过使用全局光照模型（如环境光、散射光、反射光等），可以模拟虚拟物体与真实场景的光照一致性。

（2）阴影映射和反射模型：阴影和反射是影响物体真实感的重要因素。使用动态阴影映射（如深度图阴影映射）和反射技术（如环境映射、反射映射等）可以确保虚拟物体的阴影与真实世界中的物体一致，从而避免虚拟物体与场景的脱节。

（3）材质与纹理映射：虚拟元素的材质与纹理也需要根据真实场景的环境属性进行调节。通过对虚拟物

体的纹理映射和反射贴图进行细致的处理，虚拟物体不仅在形态上与真实场景协调一致，还能够与环境光照和材质相匹配，达到视觉上的融合效果。

图 3-7 直接光照和全局光照的阴影

3.3.2 背景替换技术

背景替换技术是将前景与虚拟背景融合的技术。通常，图像分割技术是实现这一目标的基础，它通过精确提取视频中的前景部分（如人物、物体等），然后将这些前景元素与虚拟背景或环境相结合。这种技术常用于增强现实应用中，使得用户能够在现实环境中体验到虚拟信息。具体步骤如下所示：

（1）前景提取：图像分割技术通过区分前景和背景，可以对视频中的人物或物体进行精准的提取。常见的技术包括基于深度学习的分割方法（如U-Net、Mask R-CNN等），这些方法能够在复杂背景下高效地分割出前景对象。

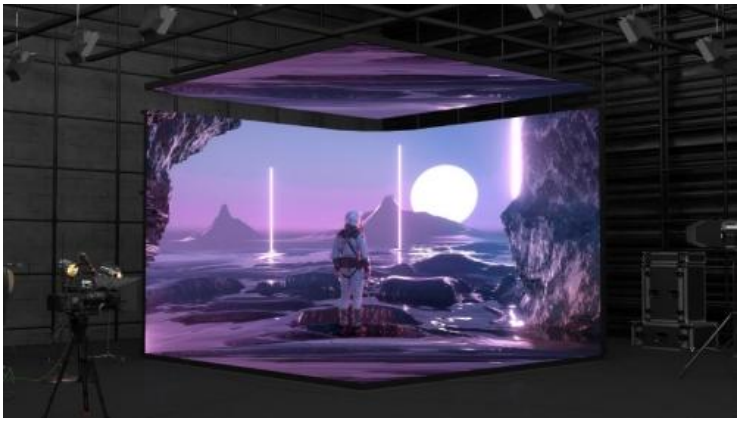

图3-8 LED虚拟拍摄系统

（2）虚拟背景合成：一旦前景被提取出来，接下来就是将其与虚拟背景进行合成。虚拟背景可以是预先设计的3D场景、动态图像，或者实时生成的虚拟环境。通过合成技术，可以在不改变前景的情况下，将其置于

完全不同的虚拟背景中，从而实现增强现实（AR）或虚拟现实（VR）中的场景重建。

（3）无缝融合：在虚拟背景合成的过程中，确保前景与背景之间的自然过渡是非常重要的。使用边缘平滑技术、颜色匹配算法和深度估计等手段，能够有效避免合成时出现的拼接痕迹，确保前景与虚拟背景的融合效果自然流畅，尤其是在动态场景中，虚拟背景的元素和前景的移动必须同步调整，保持一致性。

3.4 数实内容编码传输技术

近年来，编码传输技术经历了显著的发展，尤其是从传统视频传编码输向数实融合内容的转变。传统视频编码主要关注压缩和传输效率，而数实融合内容不仅需要高质量的视频内容，还要实现虚拟与现实的深度融合，这对编码技术提出了新的挑战。本节内容主要包括数实融合内容编码和数实融合内容传输，如图3-9所示。

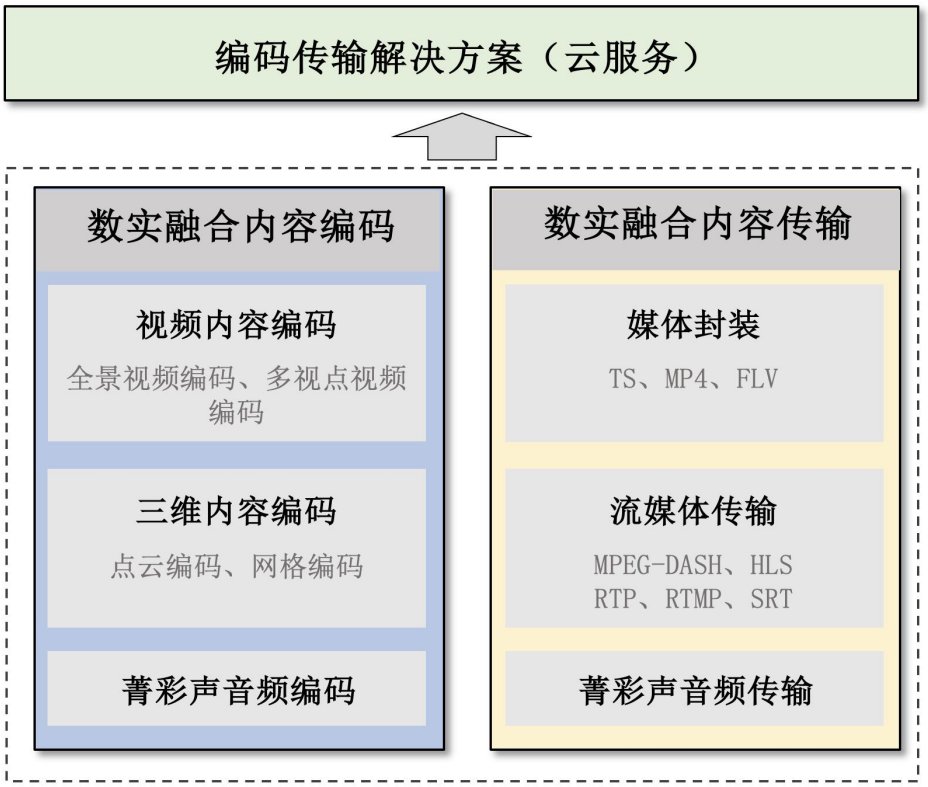

图3-9 编码传输技术架构图

3.4.1 数实融合内容编码技术

1. 全景视频编码

全景视频编码是将全景视频数据转换为适合存储和传输的数字信号的过程。由于全景视频具有高分辨率、高形变等特点，其编码过程比传统2D视频更为复杂。全景视频编码技术旨在通过有效的压缩算法，减少视频数据的存储空间，同时保持较高的视频质量。全景视频编码流程通常包括以下几个步骤：

（1）投影：将3D球形全景场景映射到2D图像上，常见的投影方法包括等距矩形投影（ERP）、立方体投影（CMP）和等角立方体投影（EAC）等。ERP是当前最常用的投影方法，它将全景场景映射为一个矩形图像，但存在边缘拉伸和形变的问题。

（2）编码：使用视频编码标准对投影后的2D图像进行压缩。常见的视频编码标准包括H.264、H.265和AV1等。这些标准通过预测、变换、量化、熵编码和环路滤波等模块，对视频数据进行高效压缩。

（3）优化：针对全景视频的特点，进行编码优化。例如，针对全景视频的不均匀采样和几何失真问题，可以采用特定的编码算法和参数设置来提高编码效率和质量。

2. 多视点视频编码

多视点视频由多视点采集系统从不同视角拍摄同一场景获得，是一种有效的数实融合视频表示方式。多视点视频具有高度的相关性，除了视点内存在帧内冗余和帧间冗余，在不同视点间还存在视点间冗余，因此需利用视差补偿等技术减少多视点视频冗余。主要技术包括视差补偿，利用视差信息建立不同视点图像中相应区域的对应关系，通过已编码的独立视点预测非独立视点[13]。

3. 点云编码

点云由若干离散点组成，每一个点包含空间三维信息以及部分属性信息（颜色，法向量，反射率等）。为了便于三维数据的组织和处理，八叉树利用二进制占用码来高效表示八叉树的结构，如图3-10所示。随着产业界对三维点云压缩需求的不断增加，MPEG下属的3D图像小组（MPEG-3DG）也逐步开展了3D点云压缩标准的制定工作。MP3DG-PCC是 MPEG-3DG早期的尝试，它提出了一个完整的点云编码帧结构和点云混合编码框架，并将传统视频编码中的IP帧结构引入到点云编码中。2017年10月，MPEG完成提案征集，并提出了三种点云编码模型。TMC1面向静态点云的编码，采用位置信息与颜色信息分离的编码方式[14]；对于位置信息，通过占用码等编码方法，设计块参数以及八叉树参数，支持点云有损和无损编码；对于颜色信息，采用空间自适应小波变换方法和游程/哥伦布编码方法进行压缩编码[15]。针对动态人物的压缩，TMC2采用基于映射的思

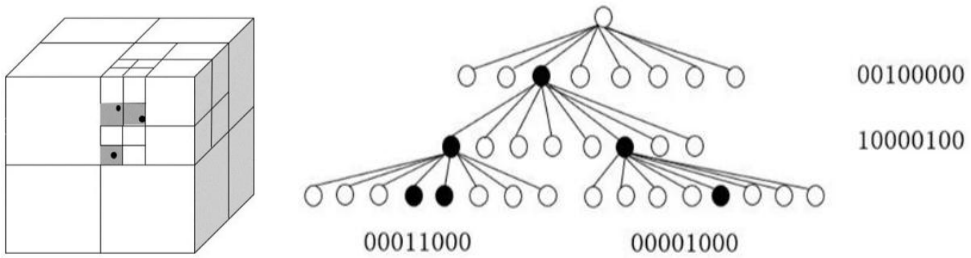

图3-10 八叉树编码

想将三维点云映射到二维平面，利用成熟的传统视频编解码器，对点云数据进行编解码[16]。TMC3针对动态物体点云，提出层次细节编码以及属性迁移概念，便于点云在网络上的传输并尽可能减小属性失真[17]。

4. 网格编码

三维网格由顶点、边和面片组成，编码物体的几何形状及属性（如纹理、颜色、法向量），广泛应用于虚拟现实、游戏和体育场景的数实融合。为高效压缩和传输网格数据，MPEG 标准下的网格编码技术采用几何、拓扑和属性分离的压缩策略。其中，MPEG-I分支提出的基于视频的动态网格编码（Video-based Dynamic Mesh Coding，简称 V-DMC）针对连接性随时间变化的动态网格进行了优化。V-DMC 将网格分解为基础网格（Base Mesh）、位移场（Displacement Field）和纹理，通过正交投影或图集打包映射到2D视频帧，利用视频编解码器（如HEVC或VVC）压缩，辅以元数据重建高质量网格。静态网格则沿用MPEG-4 Part 25的SC3DMC框架，通过量化与预测编码几何信息，采用Edgebreaker或Valence-Driven Connectivity Coding压缩拓扑，并结合游程编码处理属性。MPEG-3DG通过V-DMC和SC3DMC的结合，支持有损和无损压缩，显著降低带宽需求，适用于实时流式传输和高保真渲染。

5. 菁彩声音频编解码

在体育赛事中使用“4K+三维声”制作，结合三维菁彩声（Audio Vivid，以下简称菁彩声）实时编码传输技术，实现制作端到用户端的全流程端到端实时传输，在提升观众的沉浸感与真实感同时，拉近观众与赛场的距离。菁彩声编解码系统采用混合AI编解码架构，即预处理阶段采用传统编解码技术，在特征变换，量化熵编码阶段采用基于AI的技术，既兼顾传统音频压缩理论的精华（心理声学理论）和深度学习提取抽象特征的优势，又在算法性能和开销之间达成了合理的平衡，编码信息量也因为5.1.4+object及HOA的引入，编码信息量有巨大提升，所以在编码侧给编码器带来了较大的压力，为完整实现Audio Vivid 5.1.4+object及HOA效果，需要进行大量的编解码算法性能优化。

（1）编码算法优化

基于传统Audio vivid编码框架，对编码框架和流程进行改进，从原本的单线程串联模式，优化到多线程并行处理模式，从而实现在多核设备上的优化，重构总体编码框架，代码模块化拆分，解耦合，使各数据处理模块能够在各个时间线并行处理，并解决数据同步问题。结合Audio Vivid具体算法实现原理，在编码预处理、下混、频谱分组处理等环节研究如何实现采样点的并行处理，实现在低主频多核设备上性能的Audio Vivid编码性能提升，实现实时直播5.1.4+6的最高音质效果。

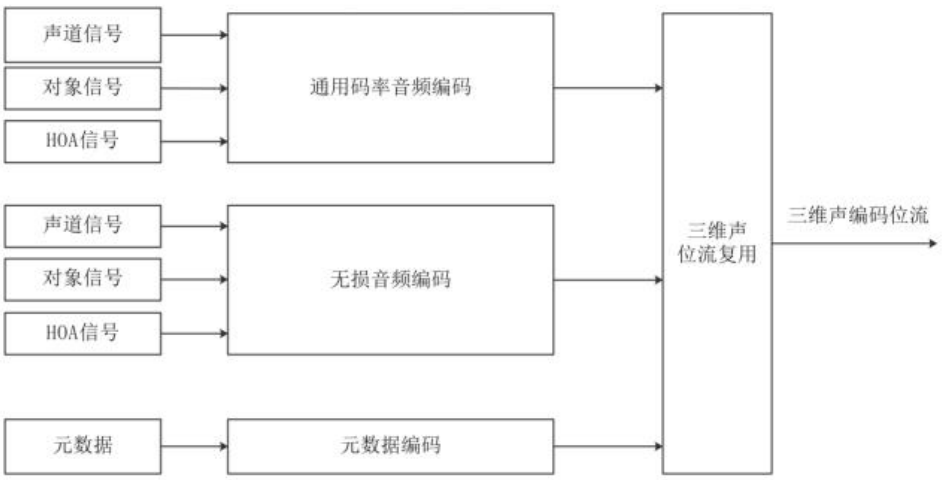

图3-12： Audio Vivid编码框架

Audio Vivid 编码框架主要分为信号预处理+编码+码流位流封装3大模块，通过对编码器架构进行重构设计，可以使这几部分实现一定程度的并行。其中信号处理部分可以分为多种信号，声道信号，HOA信号，对象信号及元数据信号，这些数据通过线程池的方式调度多核CPU，使信号预处理部分各类信号实现并行处理。同时通过预缓存一定的数据信号，把信号处理部分及后续本身串行的编码部分计算也分到多核执行，可以较大程度提高整个算法的计算并行度，充分利用CPU的多核特性。

除了对整体算法框架进行并行化改造，AI技术的引入本身极大增加了编码算法的复杂度，通过对编码算法进行逐级拆解分析，得到其复杂度热点函数，针对目前CPU的特性，对热点函数进行逐个深度优化，通过CPU的SIMD指令集，设计数据并行处理算法，减少CPU执行指令周期，通过CPU Preload指令，预加载数据到寄存器，提高Cache命中率，实现矩阵计算，DCT变换，熵编码等核心底层函数的深度优化。

（2）解码渲染算法优化

Audio Vivid在解码过程中采用了基于神经网络的变换与熵编码技术。这种技术能够更有效地处理音频信号，有效的传输带宽下提升终端听音质量；Audio Vivid支持基于声道信号、HOA（High Order Ambisonics）信号、对象信号和元数据等主流信号类型的编码。这种全面的兼容性使得Audio Vivid能够适用于多种音频应用场景，满足不同用户的需求；终端侧采用软解技术方案，对Audio Vivid播放能力机型几乎做到了全覆盖。

3.4.2 数实融合内容传输技术

随着5G、千兆光宽、Wi-Fi 6等千兆网络的建设和商用的不断深化推进人工智能、边缘云、vCDN等算力技术的加持，视频已全面融入我们的日常生活，影响着我们的连接方式、沟通形式，成为数字化进程中确定性

的基础能力。对编码视频数据传输主要涉及媒体封装和传输两大部分。

1. 媒体封装

媒体封装采用容器格式在文件中描述不同的多媒体数据元素（数据流）和元数据，并将多个数据流嵌入到媒体文件。基于不同应用场景及编码的数据特征可采用不同的封装格式，如TS，MP4，MOV，FLV，MXF，MKV等。其中，TS、MP4、FLV的具体分析见表3-2。

表3-2 具有代表性的封装协议一览

封装格式	来源	特点
TS	MPEG	是一种标准容器格式，用于对PES（Packetized Elementary Stream）包的进一步封装。它主要用来传输和（传输过程中）存储音频、视频和节目系统信息等，目的是作为规范化传输的最小单元，保证传输的可靠性，以适应不太可靠的传输。该协议扩展性比较友好，可以支持多种流媒体协议。
MP4	MPEG	其文件是一种容器格式，支持多种编码格式，使用时有很大的灵活性，并可针对业务需求和新的编码格式进行扩展，适用于不同的应用场景。MP4是高清视频存储的主流方式，主要应用在MPEG DASH 、HLS等流媒体协议中，可支持多种音视频编码类型，其fragment MP4的封装格式可支持Low Latency HLS，CMAF等超低时延的流媒体协议。
FLV	Adobe	主要用于流媒体系统，FLV包括文件头（File Header Header）和文件体（File Body）两部分，其中文件体由一系列的Tag及Tag Size对组成，可将其数据看为二进制字节流。其封装的媒体文件具有体积轻巧、封装播放简单等特点，适合网络应用。

2. 流媒体传输

目前的视频应用通常采用流媒体传输方式，其具有较强的实时性和交互性。传输下层采用UDP、TCP、IP等都是通用的以太网传输协议和标准，传输上层基于不同的应用场景、数据封装格式，采用不同的流媒体传输协议，如表3-3所示。

表3-3 具有代表性的传输协议一览

传输协议	来源	特点
MPEG-DASH	MPEG	一项基于HTTP的动态自适应流传输技术，它不限制编码格式及内容，能够根据当前带宽容量、网络性能等情况自适应地实现不同码率之间的灵活切换，在为用户提供低卡顿体验的同时保证播放内容的质量。当前，MPEG DASH协议主要应用于直播、点播等传输，以及VR视频和3D视频等。
HLS	Apple	一种基于HTTP协议的流媒体网络传输协议。HLS具有跨平台性、穿墙能力强、码率自适应、负载均衡等优点。它的工作原理是把整个流分成一个个小的基于HTTP的文件来下载。媒体播放时客户端可以选择从许多不同的备用源中以不同的速率下载同样的资源，允许流媒体会话适应不同的数据速率。

RTP	IETF	RTP协议将不同编码和封装格式的音视频数据进行再封装，加上RTP头形成RTP包进行发送。RTP协议提供抖动补偿和数据无序到达的检测机制，但并不保证传送或防止无序传送，也不确定底层网络的可靠性。RTP可为媒体数据提供实时的传输服务，如交互式音视频数据，目前市场上大多采用RTP来实时传输媒体数据。
RTMP	Adobe	一种基于TCP的协议，由多个相关协议组成的协议族。RTMP传输的数据的基本单元为Message，实际传输中的最小单元为消息块（Chunk），这样提升了数据传输时的传输速度，有效解决多媒体数据传输流的多路复用和分包的问题。RTMP协议仅需一个会话即可相互通信，具有效率高、速度快、稳定性高等特点，广泛应用于直播、视频会议、在线教育、在线游戏等实时流媒体传输。
SRT	Haivision、Wowza	一种基于UDP的协议。SRT具有安全，可靠，低延迟的优点，支持AES加密以保障端到端的视频传输安全，通过前向纠正技术保证传输的稳定性，并支持高吞吐量文件和超清视频的实时传输。

3.菁彩音视频实时编码传输

在体育赛事转播中，从节目制作端到移动端、大屏端的全流程HDR Vivid/Audio Vivid实时编码传输分发系统框架如下图3-13：

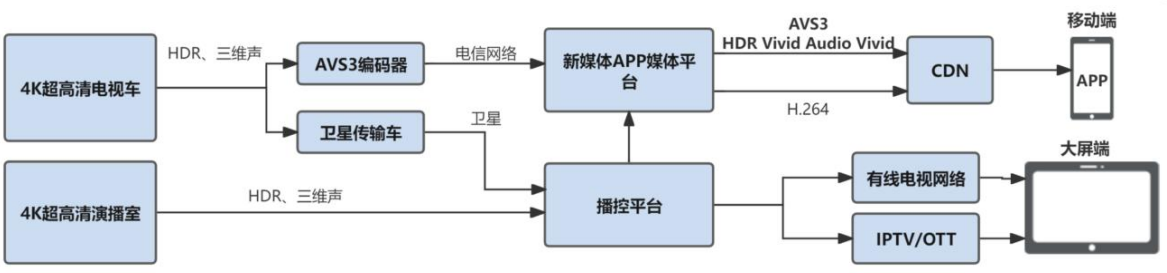

图3-13： 广播电视新媒体平台移动端应用系统框架

在制作端，使用4K超高清电视车或4K超高清演播室制作HDR+三维声节目，节目制作视频格式为3840x2160/50P/10bit/HLG/BT.2020、音频格式为5.1.4。

在编码传输部分，采用AVS3高清实时编码器，同时编码两路节目码流，一路用于移动端AVS3播放，格式为：AVS3@3Mbps/Audio Vivid@384kbps/TS@4.5Mbps，另一路采用常规高清编码方式编码，格式为：H.264@4Mbps/AAC@384kbps/TS@4.5Mbps。同时，通过卫星传输车或台内线路传送高码率信号到台内播控平台供有线电视和IPTV/OTT大屏播出使用。

在移动终端，在新媒体APP端集成了AVS3播放器，支持AVS3解码和HDR Vivid Audio Vivid实时渲染播放，节目内容通过新媒体平台进行调度分发，支持AVS3解码和HDR Vivid Audio Vivid实时渲染。

3.5 数实内容驱动与交互技术

多模态驱动与交互技术是在数实融合环境中实现自然、高效人机交互的核心支撑。通过融合多种感知模态（如视觉、听觉、触觉等）以及多维度的用户交互方式（如语音指令、手势操作、触控反馈等），该技术能够赋能沉浸式应用场景，提升用户在虚实融合环境中的参与感与沉浸感。其中，多模态驱动强调对来自多个模态（如图像、语音、文本等）的异构数据进行感知整合与深度理解，构建统一的认知与决策系统，增强系统对真实环境与虚拟内容的协同处理能力；多模态交互则聚焦于通过多种输入（如语音、动作、眼动等）与输出方式（如视觉显示、语音反馈、触觉振动等）建立双向、实时、自然的人机互动机制。这种高度融合的交互方式在沉浸式观赛、虚拟演播等数实融合应用中具有广泛的适用性和关键价值。

3.5.1 多模态驱动技术

多模态驱动（Multi-modal Driving）是指通过整合多个感知模态（如视觉、听觉、触觉等）形成统一的感知、理解和决策机制，从而驱动设备、系统或人工智能系统高效运行。这一概念主要源自对人类多模态感知与反应机制的模拟，强调了数据的多样性及其在决策中的协同作用。为了更好地理解多模态驱动的关键特征与技术挑战，可以从其核心属性进行分析。表3-4总结了多模态系统在数据整合与协同处理过程中所体现的四个核心属性，包括数据多样性、时间同步性、信息互补性和模态偏差性。这些属性不仅揭示了多模态感知机制的基本要求，也为系统设计与优化提供了理论依据和实践指导。

表3-4 多模态的核心属性

属性	定义
数据多样性	包括结构化（如传感器数据）和非结构化（如自然语言）信息。
时间同步性	模态间需要精确对齐，确保数据时间上的一致性。
信息互补性	不同模态的数据具有互补的特性，能够增强系统的鲁棒性。
模态偏差性	不同模态可能存在冗余或不一致，需要进行数据校正和优化。

单模态技术依赖于单一类型的数据（如仅依赖视觉或语音），而多模态技术通过融合多种模态，克服了单模态信息不足、数据噪声敏感等局限。同时，多模态驱动更贴近人类感知和决策方式，具备高灵活性和适应性。多模态驱动主要包括三个核心模块：模态感知、模态融合、任务决策：模态感知是通过硬件传感器或数据采集技术获取模态数据；模态融合是对不同模态的数据进行对齐与综合处理；任务决策是基于融合后的信息输出优化的任务执行策略。

（1）多模态感知技术是多模态驱动的基础，涵盖多种数据模态的采集和处理，包括视觉模态、听觉模态、触觉模态、生物信号模态，其中生物信号模态包括脑电（Electroencephalogram, EEG）、肌电（Electromyography, EMG）等。多模态感知技术的目标是解决单一模态在复杂环境下感知能力有限的问题，通过模态间的信息互补提升感知的鲁棒性和准确性。

（2）多模态融合技术是将来自不同数据模态的信息进行联合处理，以生成具有更高效能、更丰富表征能力的特征或决策结果。其目标是克服单一模态的局限性，利用模态间的冗余性和互补性，提升感知、理解和决策的准确性和鲁棒性。同时多模态融合具有很多挑战，其中包括模态特征差异，非同步性，信息冗余与噪声以及计算复杂性。

（3）多模态融合技术通常分为三种主要类别：数据级融合、特征级融合和决策级融合。数据级融合指的是直接对来自多个模态的原始数据进行整合。这种方法保持了原始信息的完整性，但需要处理高维数据，计算资源需求较高。其优点是信息完整性高，适合数据质量较高、冗余较低的场景。缺点是高维数据处理复杂，容易引入噪声。特征级融合是指通过提取不同模态的高层次特征后，进行特征对齐与整合。这种方法能减少冗余信息，同时提升融合效率。其优点是能在融合中保留模态间重要的交互特性。缺点是需要设计复杂的特征提取网络，可能丢失低层特征。决策级融合指的是每个模态独立完成预测后，将预测结果进行整合。其优点是结构简单，易于实现。其缺点是忽略模态间的深层交互，效果依赖于单个模态的预测质量。

其中，多模态融合技术近年来备受关注，大量研究工作[18-22]探讨了从理论框架、算法设计到应用实现的方方面面，尤其是深度学习方法在其中的应用与优化。其中，Campos等人[18]探讨如何将视觉模态和文本模态对齐，以提升情感分析性能，论文引入跨模态对齐技术，将图像特征与情感标签的语义特征结合。首次利用跨模态对齐提升视觉情感预测，奠定了跨模态对齐的技术基础。Radford等人[19]提出对比学习方法，通过大规模图文对齐数据训练跨模态表示模型，其方法展示了多模态表示学习的强大能力，广泛应用于图文检索、文本生成图像等任务。Baltrušaitis等人[20]系统总结多模态学习中的挑战、方法和应用，提出一种多模态学习的分类框架。他们总结的多模态在应用领域相关的技术挑战如表3-5所示。

表3-5 在应用领域相关的技术挑战[20]

应用	**挑战**				
	表示	**转换**	**对齐**	**融合**	**协同学习**
语音识别与合成					
视听语音识别	√		√	√	√
（视觉）语音合成	√	√			

事件检测					
动作分类	√			√	√
多媒体事件检测	√			√	√
情绪和情感					
识别	√		√	√	√
合成	√	√			
媒体描述					
图片描述	√	√	√		√
视频描述	√	√	√	√	√
视觉问答	√		√	√	√
媒体摘要	√	√		√	
多媒体检索					
跨模态检索	√	√	√		√
跨模态哈希	√				√

多模态驱动的任务决策技术旨在利用多个模态（例如图像、文本、语音等）的数据综合分析，以支持复杂场景下的智能决策。这些技术重点关注如何将不同模态信息有效整合，以提升决策的准确性、鲁棒性和适应性。任务决策是指通过分析输入数据，生成用于指导下一步操作或任务完成的智能化判断。传统单模态决策仅依赖单一来源的数据进行决策，例如图像分类仅依赖图像模态。而多模态决策结合多源数据（如图像与文本）共同完成复杂任务，提供更多的上下文和更高的准确性。多模态任务决策的特点有信息互补性，信息冗余性和跨模态关联。

3.5.2 多模态交互技术

多模态交互技术（Multi-modal Interaction Technologies）是现代人机交互领域的重要研究方向，通过整合和协调多种模态（如语音、视觉、手势、触觉等）数据，建立更自然、高效的交互方式。这种技术不仅突破了单一模态交互的限制，还使得机器具备更高的感知能力和更智能的响应机制。

多模态交互技术的核心在于感知、对齐、交互逻辑设计以及反馈输出。输入感知技术支持用户以多种方式与系统进行交互，包括语音、视觉、触觉，甚至脑电信号等，如表3-6所示。

表3-6 输入感知技术

类别	技术	功能	挑战
语音交互	语音识别（Automatic Speech Recognition, ASR），自然语言处理（Natural Language Processing, NLP）	通过语音指令实现设备控制或信息查询，如“播放音乐”或“打开灯”	识别噪声环境中的语音、支持多语言输入
手势交互	基于视觉的动作捕捉，如手势识别算法	通过特定手势触发动作，	如何区分无意手势

	（MediaPipe、YOLO）	例如挥手关闭屏幕	和指令性手势
眼动交互	眼动追踪传感器与Gaze Tracking算法	通过注视控制光标移动或激活某些元素	在动态界面中保持精确度和实时性
触觉交互	触摸屏、振动反馈、力反馈装置	感知用户的触摸或按压力度，并给出物理反馈	模拟真实物理触感，支持多点触控交互
生理信号交互	脑机接口（Brain-Computer Interface，BCI），肌电信号（Electromyography，EMG）识别	通过分析用户的脑电波或肌肉动作控制设备	高精度信号处理与用户意图理解

其中，最大两类包括手势识别和触控与语音控制。手势识别是通过深度传感器识别用户的手势，提供自然的交互方式。例如，用户可以通过手势操作来选择、放大或旋转虚拟物体。触控与语音控制是结合触控屏和语音识别技术，增强用户体验。用户可以通过触控屏幕与虚拟内容互动，或通过语音指令进行控制，提升交互的便捷性。

图3-9 非接触式手势识别设备

过去几十年，多模态交互技术一直是学术界和工业界广泛关注的研究方向[23-25]，催生了众多前沿技术成果。在工业应用中，非接触式手势追踪设备和体感控制系统尤为典型，代表了多模态交互在实践中的应用探索。

如图3-9所示非接触式手势识别设备，采用红外传感器和高精度的手部跟踪算法，支持640×240（红外）×2的分辨率和120帧每秒的帧率，能够实现精细的三维手势识别，这类设备广泛应用于虚拟现实（VR）、增强现实（AR）以及其他人机交互场景，为用户提供自然直观的交互方式。另一类体感控制系统则以更高的分辨率（可达1280×720（红外）×2）和约115帧每秒的性能，结合深度传感器、红外光源和RGB摄像头，实现对人体全身动作的高精度捕捉，同时支持语音识别。这些系统最初用于娱乐领域，但也迅速在工业和专业场景中得到拓展应用，进一步推动了多模态交互技术的实用化进程。

图3-10 多用户协作手势交互[29]

在学术界，Morris等人[23]探索了多用户协作手势交互，如图3-10所示，特别是在虚拟场景中的应用，为多模态交互的协作性提供支持。Turk[24]对多模态交互技术进行了全面综述，涵盖语音、手势、触觉等多种模态的应用，重点分析模态融合的技术路径以及交互系统设计中的挑战。除此之外，Jaimes等人[25]回顾了多模态交互的关键技术、实现方法和应用场景，同时分析了多模态系统在实际应用中的优势与难点。

在语音交互方面，三维菁彩声相对传统声音增加了空间感和方位感，能更好真实再现现实世界中的声音，带给观众更具感染力的临场空间感、方位感，满足人们对高质量视听的体验需求，同时可具备个性化选择和交互体验。菁彩声技术解决声音从构建到还原的整个环节，在家庭环境、影院环境、个人、AR/VR 以及车载中广泛应用，随着终端适配范围的扩大和交互功能的完善，逐渐成为体育视听体验的常态化配置。

在体育赛事转播中，利用菁彩声技术通过声音分层采集和多声道布局，可精确精准捕捉并还原赛场细节。根据拾音对象的不同，采用不同类型、指向性的多个话筒分别拾取场地声、观众声、运动员声、击球声、教练指导声、效果声等，并通过水平环绕声层与垂直顶部声道结合，构建空间声场（如图3-11），使观众能清晰分辨运动员脚步、裁判哨声、观众欢呼等不同声源的方向与距离，实现“声临其境”的体验。

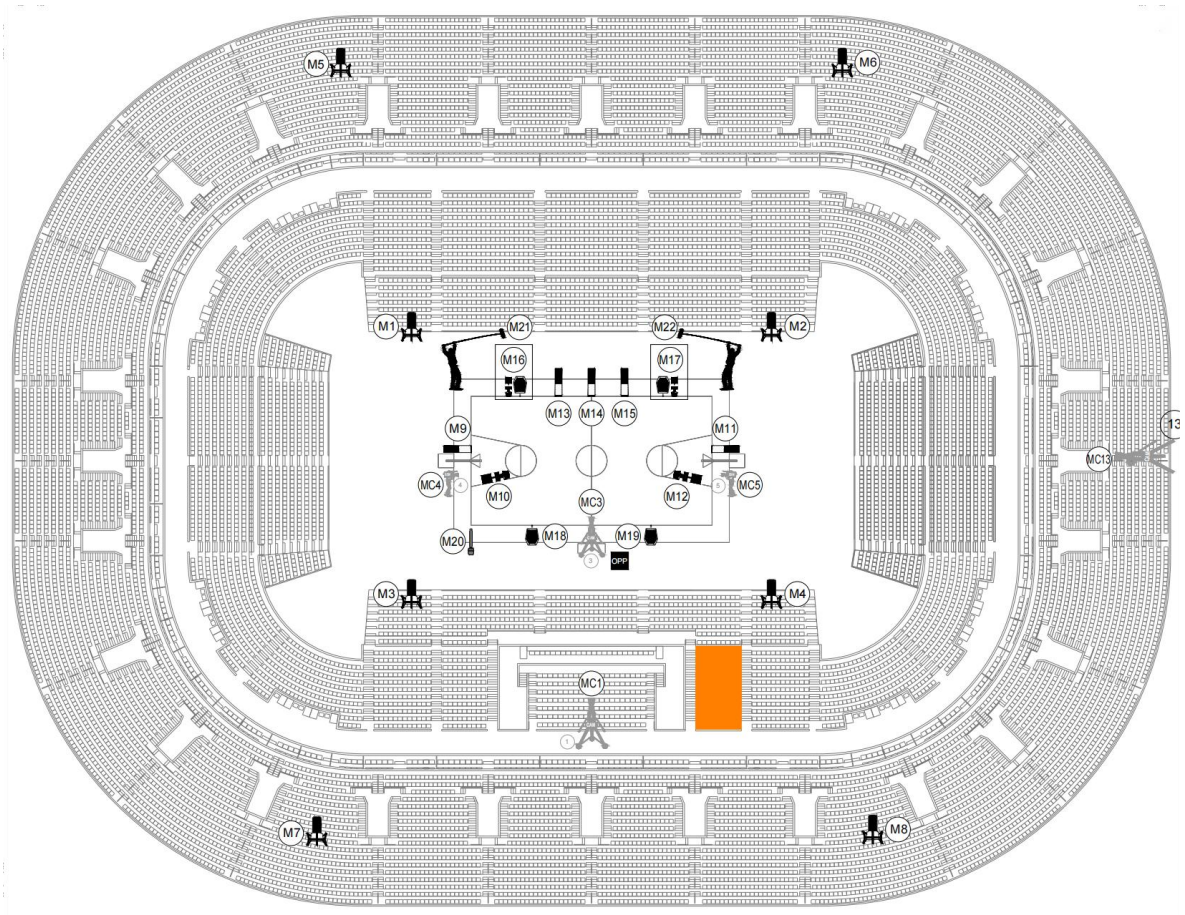

图3-11：杭州亚运会篮球项目话筒位置图

4. 数实融合系统与解决方案

4.1 支持数实融合技术的硬件设备

随着数实融合技术的不断发展，硬件设备的进步同样不可或缺。这些设备包括高分辨率摄像头、全景相机，以及专门用于增强现实（AR）和虚拟现实（VR）内容捕捉与显示的相关设备。它们在确保高质量视频输入和输出方面发挥着关键作用。

1. 采集设备

图4-1 高分辨率相机阵列

高分辨率相机阵列系统是数实融合视频系统中的核心组件，能够捕捉细致入微的图像和视频，并保持不同视角相机的时间同步。现代高分辨率摄像头通常具备4K甚至8K的分辨率，支持高帧率拍摄，能够在动态场景中保持图像清晰度。现代的全画幅无反光镜相机能广泛应用于专业视频制作和实时捕捉场景。这些设备在AR/VR内容创作中，能够提供更丰富的视觉信息，使得虚拟元素与现实环境更好地融合。

2. 显示设备

AR和VR头显设备是数实融合技术的关键展示平台，它们能够将数字内容与现实世界叠加，或完全沉浸在虚拟环境中。通过集成高分辨率显示屏、多重传感器以及强大的处理器，这些设备能够呈现高清图像，并实现对周围环境的精准感知，为用户提供更加沉浸的互动体验。这些设备不仅支持高帧率视频播放，还通过先进的眼球追踪与动作捕捉技术，实时感知用户的视线和身体动作，显著提升互动的自然性和精准性。

以图4-2中的AR头显设备为例，在工业应用和远程协作领域表现尤为出色。其52°视场角的全息波导显示屏和姿态识别传感器，可以精确识别手势并进行空间映射，用户能够更加直观地操控虚拟内容，实现人与数字世界的无缝交互。混合现实（MR）头显设备结合了AR和VR的优势，搭载超高分辨率的显示屏（单眼分辨率高达3660×3200），带来了极为清晰、逼真的视觉体验。同时，内置的计算芯片确保了低延迟的沉浸式体验。其独特的空间音频技术以及精确的手势识别能力，使用户能够通过“眼动+手势+语音”的自然交互方式与虚拟内容进行互动，而不再依赖传统控制器。这些技术突破，极大提升了虚拟与现实融合的体验，并推动了AR/VR技术在消费市场、工业应用及专业领域的广泛应用。

随着技术的不断进步，AR和VR头显设备在生产力、娱乐、远程协作等场景中的潜力日益凸显，推动了数字内容交互方式的创新，促进了虚拟与现实世界之间的无缝融合。

图4-2 AR头显设备

3. 虚拟现实视频系统

在体育赛事中通过将全景相机和高分辨率摄像头结合使用，为观众提供了沉浸式的观看体验。通过实时传输和处理，观众不仅可以在家中以VR设备感受比赛，还可以与现场观众进行互动。此外，AR技术的引入，使得赛事分析和数据可视化能够实时叠加在观众视野中，增强了信息传递的效率。

通过这些高分辨率摄像头、全景相机以及AR/VR设备的结合，数实融合技术得以充分发挥其潜力。这些硬件设备不仅提升了视频内容的质量和丰富性，也为用户提供了更加身临其境的体验，推动了数实融合技术的广泛应用。

4.2 硬件加速和优化策略在数实融合中的应用

在数实融合技术中，体育赛事等高需求场景对系统的性能和响应速度提出了严峻挑战。为确保流畅的用户体验，硬件加速和优化策略的应用变得至关重要。这些策略不仅能提升图像处理的效率，还能减少延迟，从而增强观众的沉浸感。

硬件加速是指利用专用硬件组件来执行特定任务，以提升性能。例如，在视频编码和解码过程中，使用专用的编码器（如H.264/H.265硬件编码器）可以显著提高处理速度。这些硬件编码器能够以更低的功耗和更高的效率处理高分辨率视频流，从而实现实时传输。另一个重要的硬件加速方式是使用图形处理单元（GPU）。现代GPU不仅可以用于图像渲染，还支持并行计算，适合处理复杂的图像处理算法。在体育赛事直播中，利用GPU进行图像合成和特效处理，可以确保高帧率和高清晰度的画面输出。硬件层面的优化策略还包括对数据传输路径的优化和内存管理的改进。例如，通过使用更高带宽的传输接口（如Thunderbolt或HDMI 2.1），可以提高视频数据的传输速度。此外，针对存储设备的优化，例如使用固态硬盘，可以降低数据读取延迟，确保视频流的连续性。在软件层面，流媒体传输协议的选择同样重要。采用如RTMP或SRT等高效的流媒体传输协议，可以在保证视频质量的同时，减少延迟和缓冲。利用智能算法对网络状况进行实时监测和动态调整，可以进一

步提高视频传输的稳定性。

通过硬件和软件的协同优化，可以形成一个端到端的闭环解决方案。精心设计视频捕捉、处理和传输的每个环节，确保整个系统的高效性和稳定性。软件控制系统根据实时数据反馈，动态调整视频流的编码参数和传输路径，从而实现了更高效的资源利用。利用以上硬件加速和优化策略，数实融合视频系统能够在体育赛事等高需求场景下提供流畅的用户体验。这些技术的结合，展现了未来数字媒体的无限可能，为用户带来了前所未有的沉浸式体验。

5. 业界实践与案例

5.1 国际体育赛事案例

数实融合技术，作为一种前沿的创新力量，正在多个领域展现出其深远的影响力和广泛的应用前景。如图5-1所示，数实融合技术不仅在人工智能领域中扮演着重要角色，而且在其各个技术分支中都有所体现和应用。从机器学习到自然语言处理，从计算机视觉到智能机器人技术，数实融合技术正推动着人工智能技术的进步和创新。

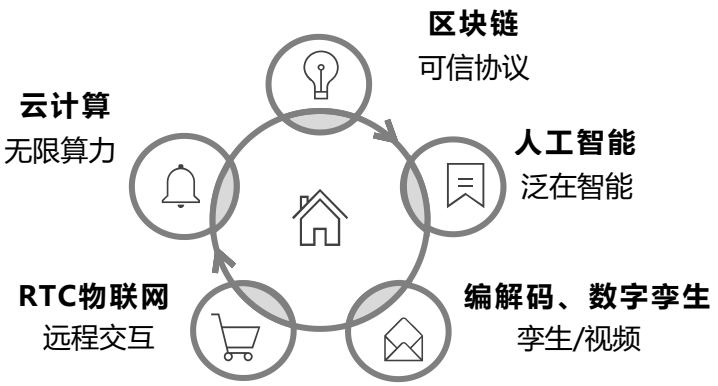

图5-1 数实融合技术在多领域的应用

此外，在体育赛事，特别是大型国际体育赛事转播中，数实融合视频技术的应用和支持尤为显著。以下是一些大型国际体育赛事转播中数实融合视频技术的应用与支持的实例：

1. 北京冬奥会

北京冬奥会开幕式及主要赛事为首次在奥运历史上采用8K超高清视频技术进行转播，提供极高的画面清晰度。通过5G网络，赛事数据实现高速、稳定的传输，如图5-2所示。结合AR（增强现实）技术，观众可体验多

视角观看等功能，大幅提升观赛互动性。依托5G与8K超高清大屏，观众能够获得沉浸式的视觉体验。这些技术应用充分体现了数实融合视频技术在体育赛事，特别是大型国际赛事转播中的关键作用，有效整合数字与现实场景，为全球观众带来更高质量的观赛体验。

图5-2 北京冬奥会现场测试图

2. 杭州亚运会

如图5-3所示，在杭州亚运会中，数实融合的视频技术得到了广泛应用，在杭州亚运会上，我国实现了全球首创“数实融合”点燃火炬创历史，为观众带来了前所未有的观赛体验。杭州亚运会首次采用4K HDR超高清视频和5.1环绕声，同时提供8K超高清电视公共信号，这标志着转播报道进入了一个新的技术时代。此外，亚运会的核心系统100%上云，这是历史上首届云上亚运会。云技术的应用不仅提高了赛事的组织和运营效率，还为观众提供了一站式的数字观赛服务。

图5-3 全球首创“数实融合”点燃火炬

3. 巴黎奥运会

2024年巴黎奥运会的技术应用包括8K超高清直播、3D全息视频技术、通过AI平台全方位分析运动数据等，

巴黎奥运会将是第一届提供端到端8K直播的奥运会，使用基于Intel Xeon处理器的服务器来编码和压缩8K信号。此外，AI还将通过数字孪生技术，如图5-4所示，通过场地的数字表现形式，优化奥运会的规划和管理，数实融合技术为观众提供了身临其境的观赛体验。

图5-4 巴黎奥运会的“AI+”裁判技术

中国移动咪咕公司的AI球星点亮黑科技，依托AI图像识别及画面智能追踪技术，在巴黎奥运期间的足球、篮球赛事中，对运动员进行实时检测、跟踪和身份识别，观众只要在直播画面上点选喜爱的球员，就能实时获取球员的统计数据，解锁游戏化观赛互动体验。

巴黎奥运会期间，中国移动咪咕公司联合北大院士团队，依托超高清元视觉技术创新实验室、超高清全媒体开放实验室两大载体，建立面向奥运的智能互动观赛系统，旨在体育赛事观赛的交互体验中创建更加真实的、具有更多自由度交互的沉浸式内容。用户可将赛场上特定运动员替换为自己的数字化身，在多维全真的数实融合沉浸式赛场中“应战”，实现基于运动员动作的人物模型驱动，增添观赛互动性和趣味性，丰富观众的参与感和体验感。

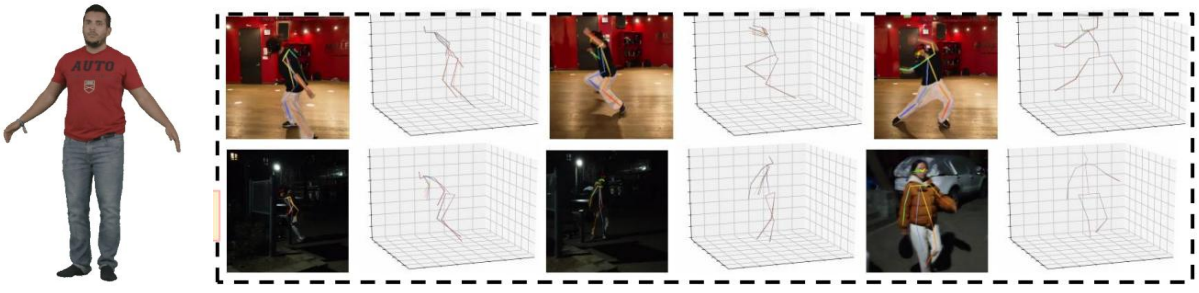

图5-5 高保真重建(左)，人体姿态估计(右)

在内容构建技术方面，如图 5-5（左），该系统采用低成本高保真数字人重建技术，仅用单目视频即可完成人物模型重建；同时，在神经场加速结构和动态场景空间跳跃策略技术的加持下，可在几分钟构建观众数字化身，并兼顾重建质量和姿态合成效果。如图 5-5（右），通过三维人体姿态估计技术，结合基于重投影的假设选择网络和时域自适应提升技术，系统能够细粒度地感知运动员的空间位置和运动姿态，精确追踪他们在赛场上的运动轨迹。从运动员的空间位置到运动姿态，系统都能全面捕捉，为将数字化身无缝融入赛场环境提供了坚实的基础。

图5-6 基于人体先验的动态分层数字人编码

在编码传输技术方面，该系统基于人体先验的动态分层数字人编码：在编码端学习全局结构信息，提取每帧高精度位姿信息作为动作层，同时引入法线图保留细粒度的几何细节，设计分层紧致表示模型并独立压缩；在解码端以生成的方式重建动态三维人体序列。实现三维数字人高效压缩，确保高质量内容的流畅传输。

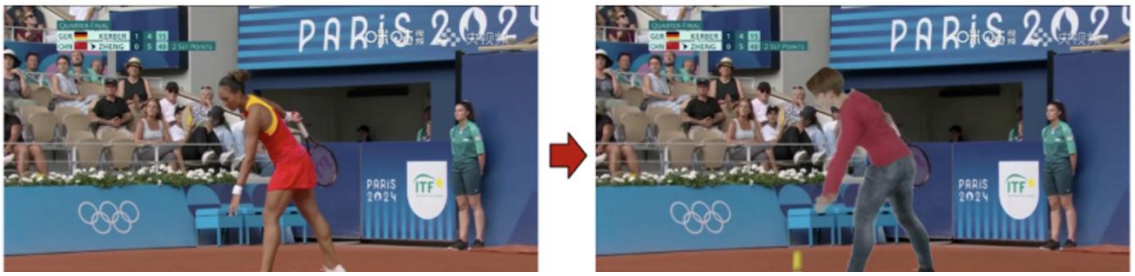

图5-7 观众可将网球运动员的形象替换为自己

在驱动交互技术方面，该系统采用了创新的分离式设计，让观众的重建端与运动员的姿态捕捉端各自独立，再进行数实融合。如图5-7，可以轻松地将不同观众模型替换到任意奥运赛事中，或者在同一场比赛中体验多种角色转换，成为赛事主角，获得更加沉浸的体验。

5.2 国内体育赛事案例

2024年11月2日，2024赛季中国足球超级联赛最后一轮，上海海港坐镇主场浦东足球场对阵天津津门虎的收官之战见证了亚洲足球转播史上的里程碑时刻：中国移动咪咕公司联合百度智能云、索尼中国以及Origem Sports（实刻体育科技）公司，进行了一次数实融合转播技术的联合测试，如图5-8所示，完成了从设备部署、网络传输、数据采集、数据加工、多角度可视化到移动端咪咕视频App内多线路观看的全链路、端到端实时转播能力验证。

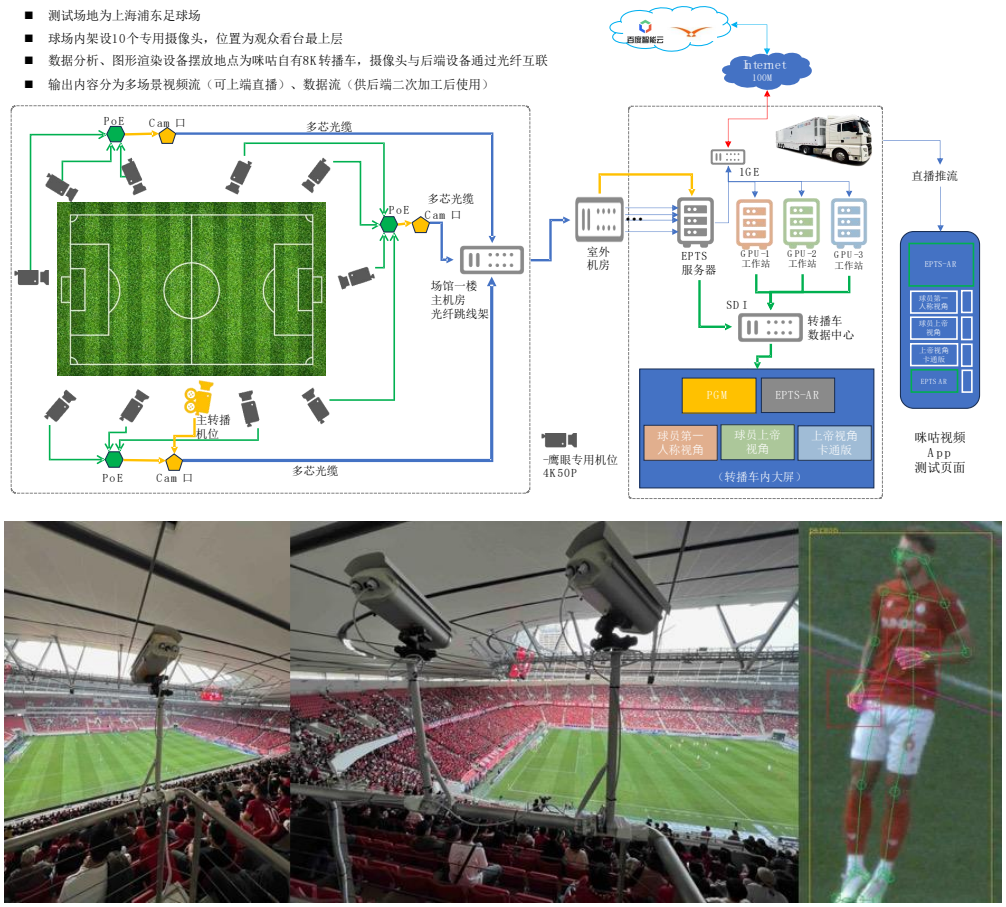

图5-8 国内体育赛事转播实例

以无穿戴非接触式的光学骨骼追踪系统为硬件基础，通过EPTS（Electronic Performance Tracking System 电子运动员表现追踪系统），在软件层面融合云端骨骼姿态算法及消息队列，完成数据实时采集、实时3D骨骼姿态拟合、超实时逆向解IK、人体映射动作驱动、3D背景重建、虚拟摄像机视角切换、3D可视化交互、直播推流等一系列复杂技术流程的串联，最终在延迟时间小于30秒的情况下，实现了转播主机位AR（现实增强）数字特效、3D自由视角等创新转播应用。

在边缘计算强大的算力支持下，实时采集的骨骼点位数据经过格式整理、数据科学运算以及场景化深度学习后，可以生成多维度、更全面、更精准的数据展示，结合场上球员的球衣号码OCR识别和坐标位置匹配技术，无需切换就能让屏幕前的观众体验沉浸式观赛，如图5-9，降低传统体育观赛理解门槛的同时，也通过放大和丰富比赛细节，在赛事直播过程中赋予观众更多观赛乐趣。

图5-9 沉浸式观赛时的射门球速呈现

本次测试还有另一个创新点，即“重造第二现场”：如图5-10绑定咪咕数字人形象的实时比赛场景重构。既放大了现实体育比赛IP的现有价值（即在直播过程中为观众带来电竞转播中惯用的“观察者视角”全新转播体验），又能为IP采购方创造更丰富的额外商业价值（比如，通过前置的数字资产可以把更丰富的企业品牌形象在3D空间中进行伴随展示和露出）。此创新点论证了数据采集技术和边缘计算技术可以赋能企业品牌形象和赞助权益的更多激活与传播可能性。

图5-10 实时比赛场景重构

此外，鉴于测试案例中所获取的比赛数据，相较传统采集方式来得更丰富、更全面、更精准，数据经过加工处理，还能产生更多的使用与消费场景。如：辅助训练、短视频自动生产、球迷竞猜互动等更多元的赛事应用场景。真正通过科技，向世界讲述全新的中国故事。

6. 技术产业发展趋势

6.1 数实融合技术的发展

数实融合技术在未来展望中呈现出无限的可能性和广阔的前景。随着技术的不断进步和创新，特别是在5G、物联网、人工智能等领域的快速发展，数实融合将进一步深化，为新质生产力提供更为强大的动力。如图6-1所示，在未来应用中，数实融合将使生产过程更加智能化、自动化，大幅提高生产效率和质量，同时促进产业结构的进一步优化升级。数实融合技术的应用还涉及到数字孪生技术，即物理实体在数字世界的孪生，强调数字世界与物理世界的一致性。数字孪生技术能够构建物理实体的虚拟副本，而AR技术则能够将虚拟信息叠加到真实世界中，两者结合为用户提供了沉浸式的体验，并在教育、培训等领域提供直观、生动的学习环境[26]。

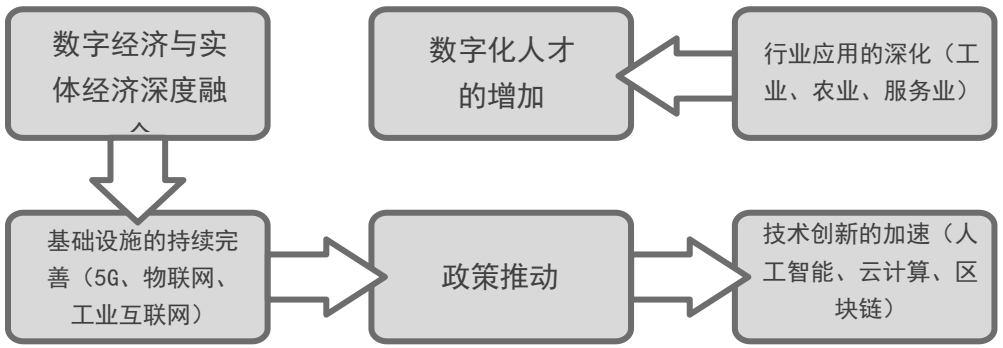

图6-1 数实融合技术的不断推进

1. 硬件性能进展

硬件性能的提升是数实融合技术发展的重要基础。随着技术的进步，通信网络基础设施不断完善、算力基础设施不断增强、硬件设施逐步实现标准化和互联互通、智能硬件产业不断发展、数据采集和边缘计算的能力逐渐提升、云原生开发体系成功构建、智能应用创新技术系列不断融合。如8K超高清显示技术、5G传输技术、AI芯片等为视频技术提供了强大的支持。未来，硬件性能的持续提升将进一步推动视频技术的革新。

2. 软件配套发展

软件配套方面，工业数字化能力不断提升、工业互联网基础设施不断完善、智能制造示范工厂生产效率不

断提升、大数据软硬件和服务方案市场不断拓展、系统解决方案市场增长、智能技术集成融合、各种数实融合的技术支持增加。如云技术、AI算法、大数据分析等，为视频内容的生产、处理和分发提供了高效的工具。软件的发展使得视频内容的个性化推荐、智能编辑成为可能。

3. 视频内容技术革新

视频内容技术方面，如数字孪生、3D全息、VR/AR等技术的应用，为观众提供了更加沉浸式的观看体验，这些技术的应用不仅提升了赛事的观赏性，也为未来体育赛事的转播和传播提供了新的方向和可能。

6.2 未来发展趋势与规划

尽管数实融合技术在多个方面取得了进步，但仍存在一些挑战，如数据安全和隐私保护、技术标准化、跨平台兼容性等问题[27]。这些问题需要行业共同努力，通过技术创新和政策引导来解决。此外，提升用户体验是数实融合技术发展的核心目标之一。未来，通过进一步优化视频质量、降低延迟、提供个性化服务等，可以进一步提升用户体验。成本控制也是数实融合技术发展的关键问题，随着技术的发展，需要平衡技术创新与成本之间的关系，通过技术优化和规模化应用来降低成本。

行业政策在推动数实融合技术发展中起着重要作用。政策的支持可以为技术研发、产业升级提供指导和保障。未来，应继续加强政策支持，推动产业链现代化，促进数实融合技术的健康发展。表6-1基于技术规划、企业规划、政策规划给出了一些产业链环节的未来规划建议。

表6-1 基于数实融合技术展望的未来规划

类别	未来规划建议
技术规划	1、加强关键技术攻关，加强集成电路、人工智能、关键软件、云计算等技术的研发 2、促进数实融合深度赋能千行百业，准确把握数字化、网络化、智能化发展趋势 3、促进要素转型，推动数据要素化 4、夯实基础底座，加强5G网络、数据中心、工业互联网等数字设施建设
企业规划	1、加快企业转型 2、深化行业赋能 3、推动企业园区升级 4、构建产业公共数据空间
政策规划	1、深化公共数据，加快数据立法、构建开放生态 2、推动开放创新，实施动态开放 3、加强政策保障，完善顶层规划设计，优化新型基础设施的管理措施

7. 政策与标准化建议

7.1 政策背景

随着数字技术的飞速发展，数实融合技术正逐渐成为推动各行各业数字化转型的重要力量。然而，在应用过程中，这一技术也面临着诸多挑战，同时也孕育着巨大的机遇。数实融合技术涉及VR、AR、点云等核心技术的深度整合，技术门槛高，融合难度大。此外，随着5G、人工智能等技术的不断革新，人机交互界面正在迭代，如何将这些新技术与数实融合技术有效结合，实现技术创新的突破，是当前面临的一大挑战。在数实融合视频技术的应用中，数据流通是关键。然而，当前产业链中存在数据交易机制不成熟、产业数据资源归集和利用水平低、公共数据开放不充分等问题，导致数据孤岛和数据垄断现象频发，制约了数据要素对企业生产活动的乘数效应。同时，数据在共享、传输、处理中的安全隐患也不容忽视。数实融合技术旨在提供沉浸式、交互式的用户体验。如何实现高质量的HDR、3D视觉和虚拟环境适配，以及多自由度的交互体验，是当前技术发展的难点。此外，高昂的技术研发和应用成本也是制约数实融合技术广泛应用的重要因素。

近年来，我国制定了一系列支持数字经济高质量发展的政策，积极推进数字产业化、产业数字化，为数实融合视频技术的发展提供了良好的政策环境[28-30]。同时，随着5G、人工智能等技术的普及，各行各业对数实融合视频技术的需求不断增长，特别是在体育赛事观赛体验、智慧体育等领域，数实融合技术有着广阔的应用前景。随着VR、AR、点云等核心技术的不断突破，数实融合技术能够实现更深度地沉浸式体验，主要体现在真实场景的六自由度、更清晰和流畅的内容显示、多通道交互等方面。这些政策的制定将推动相关产业的升级和转型，促进新业态、新模式的涌现。

7.2 标准化建议

随着数字经济与实体经济的深度融合，数实融合相关标准在推动产业数字化转型、促进数据流通与共享、增强体育赛事体验等方面发挥着至关重要的作用。以下是国内外在数实融合领域的关键标准与组织：

一、国内核心标准与组织

1. 音视频编解码标准（AVS）

AVS系列标准：AVS（Audio Video coding Standard）是我国自主制定的音视频编解码标准，旨在为数字音视频产业提供高效、先进的编解码技术。AVS标准涵盖了视频编码、音频编码、系统、测试等多个方面，广泛应用于广播电视、视频监控、视频会议、移动多媒体等领域。AVS标准的不断演进和推广，有助于提升我国在音视频领域的技术自主性和产业竞争力，为数实融合中的多媒体数据处理和传输提供了坚实的技术支撑。

AVS3：作为AVS系列标准的最新版本，AVS3在编码效率和性能上有了显著提升，能够更好地适应高分辨率、高帧率、高动态范围等视频内容的编解码需求，为超高清视频产业的发展提供了有力支持，也为数实融合场景下的沉浸式体验、智能媒体应用等提供了高效的数据压缩和传输解决方案。

数实融合相关标准：数实融合领域最重要的部分是数字人技术，其主要涵盖了图像、视频、三维数据、语音等多模态信息，每种模态的编码方式、格式及传输效率直接影响呈现效果。由于缺乏统一标准，不同平台与应用间兼容性差，数据在编码系统间转换时易出现丢失或畸变。为解决数字人领域标准缺失问题，AVS标准工作组正在开展探索三维数字人的高效表示和编码技术规范，2024年12月，AVS工作组成立了三维数字人高效表示和编码标准联合推进组。2025年3月，由AVS需求组、视频组、点云组和VRU组联合决议，建议由AVS VRU组、AVS点云组分别设立数字人编码探索小组，从三维高斯表征以及点云数据角度探索数字人相关标准技术研究和标准制定推进工作。以期为数字人压缩表示提供标准方案，进一步服务数实融合内容表达。

2. 电子标准院（CESI）

中国电子技术标准化研究院（电子标准院，China Electronics Standardization Institute）：作为我国电子信息技术领域的权威标准化研究机构，电子标准院在数实融合相关标准的制定、推广和应用方面发挥着重要作用。电子标准院参与了多项国家和行业标准的制定工作，涵盖了电子信息产品、信息技术服务、工业互联网、智能制造等多个领域，为数实融合的标准化工作提供了全面的技术支持和服务保障。

智能制造相关标准：电子标准院积极推动智能制造领域的标准化工作，制定了包括智能制造能力成熟度模型、智能制造系统架构、工业互联网平台等在内的多项标准，为制造业的数字化、智能化转型提供了明确的指导和规范，促进了制造业与信息技术的深度融合，推动了数实融合在制造业中的广泛应用。

二、国际核心标准与组织

1、ISO/IEC MPEG

ISO/IEC MPEG（Moving Picture Experts Group）：MPEG是国际标准化组织（ISO）和国际电工委员会（IEC）联合成立的专家组，负责制定音视频及相关多媒体数据的编码标准。MPEG系列标准（如MPEG-1、MPEG-2、MPEG-4、H.264/AVC、H.265/HEVC等）在全球范围内得到了广泛应用，为数字音视频产业的发展奠定了坚实的基础。在数实融合的场景中，MPEG标准不仅保障了音视频数据的有效压缩和传输，还为多媒体内容的创作、分发和消费提供了标准化的框架，推动了多媒体技术在各个领域的深度融合和创新发展。

2、ITU-T VCEG

ITU-T VCEG（Video Coding Experts Group）：VCEG是国际电信联盟电信标准化部门（ITU-T）的研究组，专注于视频编码技术的研究和标准制定。VCEG的工作成果对视频通信和视频流媒体等领域的发展产生了深远影响，其制定的视频编码标准（如H.264/AVC、H.265/HEVC等）在提高视频传输效率、降低带宽需求方面发挥了重要作用。在数实融合的背景下，VCEG的标准为视频数据的高效处理和传输提供了关键技术支持，促进了视频技术在远程医疗、在线教育、智能交通等领域的广泛应用。

三、垂直领域技术标准

1、3D视频标准

3D视频标准：随着3D显示技术的不断发展，3D视频标准也在逐步完善。例如，MPEG组织制定了MPEG-C Part 3标准，用于支持3D视频的编码和传输；ITU-T也开展了相关工作，推动3D视频技术的标准化进程。这些标准为3D视频内容的制作、分发和播放提供了规范和指导，有助于推动3D视频技术在娱乐、教育、工业设计等领域的应用，为用户带来更加沉浸式的视觉体验，促进了视频技术与现实场景的深度融合。

2、点云压缩（PCC）

点云压缩标准：点云数据作为一种重要的三维数据表示形式，在虚拟现实、增强现实、自动驾驶、工业测量等领域有着广泛的应用。为了有效压缩和传输点云数据，国际上制定了一系列点云压缩标准，如MPEG Point Cloud Compression（PCC）标准。这些标准通过优化点云数据的编码方式，提高了点云数据的存储和传输效率，为3D数据在数实融合场景中的应用提供了有力支持，使得3D内容能够更便捷地融入到各种数字化应用中。

数实融合的推进离不开标准化的引领和支撑。上述标准和组织在各自的领域内发挥着重要作用，共同推动着数字技术与实体经济的深度融合。然而，数实融合是一个复杂而动态的过程，涉及到多个领域的交叉和协同。因此，未来还需要进一步加强标准之间的协调与配合，形成更加完善的数实融合标准体系，以更好地满足产业发展的需求，促进数字经济与实体经济的高质量融合发展。

为了继续推动数实融合技术的持续创新与发展，对相关产业联盟、标准组织提出以下建议：

（1）加强技术研发与创新，聚焦底层基础技术和核心关键技术，探索研发机构在核心关键技术攻关的新路径。同时，加强数字技术创新领域的知识产权保护力度，完善相关法律法规和政策保障体系。

（2）推动产业数字化转型，在工业制造、体育赛事等领域打造数字化转型全产业链生态系统，引导上下游企业通过“以数换数”新模式实现产业数据的互通。同时，鼓励企业运用数字技术优化生产流程、提高生产

效率和质量。

（3）加强产业间的合作与交流，构建数实融合生态系统。通过开放创新、合作共享、平等参与的全新合作模式，推动技术的持续创新与发展。同时，加强与国际先进技术的交流与合作，引进和消化吸收国际先进技术成果。

（4）制定和完善技术标准，针对数实融合技术的特点和应用需求，制定和完善相关技术标准和系统规范。特别是要关注HDR、3D视觉和虚拟环境适配到数实融合中的技术标准和规范以及多自由度交互体验的实现方式和技术要求。这将有助于推动技术的规范化、标准化发展，提高技术的通用性和可移植性。

（5）加强国际标准化合作，建议加大与国际标准化组织的合作，特别是与国际电信联盟（ITU）、第三代合作伙伴计划（3GPP）等组织的合作。ITU作为全球信息通信领域的顶级标准化机构，其在数实融合领域的标准化工作至关重要。3GPP则在通信网络的标准化中发挥着关键作用，特别是在5G及其未来发展方向对数实融合技术的推动中起着至关重要的作用。此外，中国移动集团与北京大学等科研机构的积极参与，对于引导和推动国际标准的制定具有重要作用。通过与国际标准化组织的对接与合作，可以将我国的技术优势和创新成果融入国际标准体系，从而提升我国在全球标准制定中的话语权，促进全球技术的互联互通和兼容性。

（6）推动国内标准化进程，建议进一步推动和完善与数实融合技术相关的标准体系，尤其是AVS（音视频编解码标准）、UWA（超高清标准）、SVAC（网络视频监控标准）、CUWA（云端无线视频标准）等国内标准组织的支持。这些国内标准在音视频编解码、无线通信、视频监控等领域取得了显著进展，未来应将其技术成果与数实融合的实际需求结合起来，推动跨领域的技术标准化。通过加大国内标准的研发与实施力度，可以有效促进技术的一致性与兼容性，为数实融合的广泛应用奠定坚实的技术基础。

8. 结论与展望

随着数字化技术的快速发展，数实融合技术在体育场景中的应用已经展现出巨大的潜力和价值。通过虚拟内容生成、三维建模、视频处理与编码以及数据传输等技术的结合，不仅提升了体育赛事的观赏性和互动性，还为观众提供了更加丰富和沉浸式的观赛体验。特别是在大型国际体育赛事中，数实融合技术的应用已经成为提升赛事品质、吸引观众关注和增强赛事影响力的重要手段。

然而，数实融合技术的发展仍面临诸多挑战。一方面，技术门槛高、融合难度大，需要持续投入研发力量

进行技术创新和突破。另一方面，数据流通、隐私保护、技术标准化等问题也需要行业共同努力，通过加强合作与交流，推动技术的规范化、标准化发展。

展望未来，数实融合技术将在体育场景中发挥更加重要的作用。随着5G、人工智能等新技术的不断革新，人机交互界面将不断迭代，数实融合技术将与这些新技术深度融合，实现更加智能化、个性化的观赛体验。同时，随着技术的不断成熟和成本的降低，数实融合技术将在更多领域得到应用和推广，为各行业数字化转型提供有力支持。为了推动数实融合技术的持续创新与发展，需要加强技术研发与创新，聚焦底层基础技术和核心关键技术攻关。同时，推动产业数字化转型，构建数实融合生态系统，加强产业间的合作与交流。此外，还需要完善相关法律法规和政策保障体系，为数实融合技术的发展提供良好的政策环境。通过这些努力，相信数实融合技术将在未来发挥更加重要的作用，为体育赛事和各行各业的数字化转型注入新的活力。

9. 附录

9.1 缩略语

下列术语和定义适用于本文件：

尺度不变特征变换（Scale-invariant feature transform，SIFT）

等角立方体投影（Equi-Angular Cubemap projection，EAC）

等距矩形投影（EquiRectangular Projection，ERP）

多视图立体视觉（Multi-view stereo，MVS）

高动态范围（High Dynamic Range，HDR）

混合现实 （Mixed Reality，MR）

肌电（Electromyography，EMG）

加速稳健特征（Speeded Up Robust Features，SURF）

结构运动恢复（Structure from Motion， SFM）

基于区域的卷积神经网络（Region-CNN，R-CNN）

立方体投影格式（CubeMap Projection，CMP）

脑电（Electroencephalogram，EEG）

脑机接口（Brain-Computer Interface，CI）

你只看一次（You Only Look Once，YOLO）

人工智能（Artificial Intelligence，AI）

生成对抗网络（Generative Adversarial Network，GAN）

通过检测进行跟踪（Tracking by Detection，TBD）

图形处理单元（Graphics Processing Unit，GPU）

虚拟现实 （Virtual Reality，VR）

英特尔至强处理器（Intel Xeon）

用于生物医学图像分割的卷积神经网络（Convolutional Networks for Biomedical Image Segmentation，U-Net）

语音识别（Automatic Speech Recognition，ASR）

运动图像专家组（Moving Picture Experts Group，MPEG）

增强现实 （Argument Reality，AR）

自然语言处理（Natural Language Processing，NLP）

9.2 引用

[1] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[2] Kingma D P. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.

[3] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[J]. Advances in neural information processing systems, 2020, 33: 6840-6851.

[4] 薛子育,刘庆同,周康能.视频内容生成方法研究综述（上）[J].影视制作,2024,30(02):42-47.

[5] Ho J, Salimans T, Gritsenko A, et al. Video diffusion models[J]. Advances in Neural Information Processing Systems, 2022, 35: 8633-8646.

[6] 王鹏.文本到图像生成方法的研究进展[J].信息技术,2024,(07):148-159.DOI:10.13274/j.cnki.hdzj.2024.07.024.

[7] 龚雨楠,薄一航.AIGC技术创建三维数字内容的研究进展及应用浅析[J].现代电影技术,2024,(02):26-34.

[8] 徐松.三维动画设计中的实时渲染技术应用[J].电视技术,2024,48(08):79-81.DOI:10.16280/j.videoe.2024.08.022.

[9] J. L. Schönberger and J. -M. Frahm. Structure-from-Motion Revisited, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 4104-4113, doi: 10.1109/CVPR.2016.445.

[10] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., & Szeliski, R. (n.d.). A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR' 06) (Vol. 1, pp. 519–528). IEEE. https://doi.org/10.1109/CVPR.2006.19

[11] Mildenhall B, Srinivasan P P, Tancik M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2021, 65(1): 99-106.

[12] Kerbl B, Kopanas G, Leimkühler T, et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering[J]. ACM Trans. Graph., 2023, 42(4): 139:1-139:14.

[13] Lukacs M. Predictive Coding of Multi-viewpoint Image Sets, IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), pp. 521-524, 1986.

[14] N17340 Point Cloud Compression Test Model for Category 1 v1 ISO/JCT SC29 WG11 Gwangju Jan. 2018

[15] Malvar. Adaptive Run-Length/Golomb-Rice Encoding of Quantized Generalized Gaussian Sources with Unknown Statistics, Data Compression Conference (DCC), pp. 23-32, 2006.

[16] N17348 Point Cloud Compression Test Model Category 2 v1 ISO/JCT SC29 WG11 Gwangju Jan. 2018

[17] N17349 Point Cloud Compression Test Model Category 3 v1 ISO/JCT SC29 WG11 Gwangju Jan. 2018

[18] Campos V, Jou B, Giro-i-Nieto X. From pixels to sentiment: Fine-tuning CNNs for visual sentiment prediction[J]. Image and Vision Computing, 2017, 65: 15-22.

[19] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//International conference on machine learning. PMLR, 2021: 8748-8763.

[20] Baltrušaitis T, Ahuja C, Morency L P. Multimodal machine learning: A survey and taxonomy[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(2): 423-443.

[21] Tsai Y H H, Bai S, Liang P P, et al. Multimodal transformer for unaligned multimodal language sequences[C]//Proceedings of the conference. Association for computational linguistics. Meeting. NIH Public Access, 2019, 2019: 6558.

[22] Chen M, Radford A, Child R, et al. Generative pretraining from pixels[C]//International conference on machine learning. PMLR, 2020: 1691-1703.

[23] Morris M R, Huang A, Paepcke A, et al. Cooperative gestures: multi-user gestural interactions for co-located groupware[C]//Proceedings of the SIGCHI conference on Human Factors in computing systems. 2006: 1201-1210.

[24] Turk M. Multimodal interaction: A review[J]. Pattern recognition letters, 2014, 36: 189-195.

[25] Jaimes A, Sebe N. Multimodal human–computer interaction: A survey[J]. Computer vision and image understanding, 2007, 108(1-2): 116-134.

[26] 21世纪经济报道数字经济课题组. (2022). 《中国数实融合发展趋势白皮书 第2部分：数实融合技术基本设施》[报告]，2022. [6-1]

[27]PAN Jiaofeng, WU Jing. Integration of digital and real economies to shape new advantages in development: New form of human-cyber-physical ternary fusion. Bulletin of Chinese Academy of Sciences, 2024, 39(6): 1012-1021 [6-2]

[28] “十四五”数字经济发展规划. 中国政府网，中华人民共和国国务院. 2021.

[29] 虚拟现实与行业应用融合发展行动计划（2022—2026年）. 中国政府网，中华人民共和国国务院. 2022.

[30] 周闻韬. 持续推进数字技术和实体经济深度融合. 新华社，2023年9月6日.

UHD World Association
世界超高清视频产业联盟

联系我们：
UWA联盟邮箱：support@theuwa.com
UWA联盟官网：www.theuwa.com