基于计算机视觉方法的大规模相机网络系统研究与实现

1 绪论

视觉媒体是一种十分重要的媒体内容。伴随着传感器技术、通信技术以及计算机技术的进步，用于采集图像或视频的视觉系统也从传统的单一相机系统发展为复杂的智能相机网络系统。图像或视频等媒体形式在医疗、娱乐和教育等领域都具有重要的应用价值。

在医疗、娱乐和教育等领域，出于各种目的，经常需要对动态事件进行记录和回放。通过摄影机等设备直接记录视频然后回放的传统方式能够满足基本的记录和回放要求，并且视频的精细度也随着图像传感器的进步而不断提高。然而，这种传统方式很难为用户提供足够友好的沉浸式体验。记录下的视频尽限于有限的几个相机机位，用户在回放时也只能从有限的角度观看，并且基本无法在回放时与用户进行互动。

非侵入式，无标签的

国内外研究现状

相机标定

三维重建

现实生活中的各类事件和场景中往往存在有大量人物，因而人体外形重建和动作追踪一直是计算机视觉技术的重要应用领域。如前文所述，现有的许多动作捕捉系统往往需要演员佩戴特殊标签，并在有限的演播室环境中活动。现有的计算机视觉技术已经能够满足无标签的动作追踪。针对这类特定的应用，基于模型的重建方法往往能取得较好的效果。Strack和Hilton[5]描述了一种将骨骼蒙皮与多个图像中的轮廓进行匹配的技术。Carranza等[6]将一个通用人体模型与多幅图像的轮廓进行匹配，并根据图像中具体人物的轮廓对该通用人体模型进行适配从而得到期望的重建结果。Theobalt等[7]

多相机系统

随着计算机处理能力的不断提高，许多多相机系统已经能够通过算法的优化实现实时场景重建。现有的快速场景重建方法主要分为两大类。其中一类方法基于体素离散化技术，另一类方法主要基于空间多面体相交技术。

通过体素离散化技术，Cheung G.等实现了较高精度的实时场景重建，重建场景中单个体素达到了3cm x 3cm x 3cm。Grau O.等则在相关系统中进行低精度的实时重建，并在此基础上实现了头部追踪，用于与用户进行一定的交互。该系统也可以以离线方式进行高精度的场景重建。Hasenfratz J.-M.等使用4部相机实现了场景的实时重建和渲染，在不考虑渲染性能的前提下，对单帧场景的重建仅需要小于25ms的延迟，同时该系统也具备交互功能。

通过空间多面体相交技术，Matusik W.等使用4部相机实现了实时场景重建，并在渲染时基于优化的非结构化光照图渲染算法（unstructured lumigraph rendering, ULR）[]为场景赋予了贴图。Allard J.等同样基于空间多面体相交技术得到了类似的结果。

结构光方法也是一类快速场景重建方法。Waschbusch M.等借助多组模块化的结构光单元实现了高精度实时场景重建，每组结构光单元都包含了三部相机、一部投影仪以及对应的工作站。该系统具有较好的可扩展性，结构光单元之间不会互相影响，多个单元的加入不会增加场景重建的耗时。然而，这类方法由于需要借助投影仪进行场景重建，因此对场景具有一定的侵入性。

对比而言，离线式系统往往能够完成更高精度的场景重建工作。

其它应用

Starck J, Maki A, Nobuhara S, et al. The multiple-camera 3-D production studio[J]. Circuits and Systems for Video Technology, IEEE Transactions on, 2009, 19(6): 856-869.