

第一章 全景视频生成技术

“全景”的概念最早是在欧洲艺术领域提出的，随后工程师们渐渐发现“全景”的优点和发展前景，全景应用由艺术领域转向了工程。全景视频，也称为 360 度视频，在近年来随着虚拟现实（VR）技术的快速发展而兴起，成为未来视频服务的新型载体。全景视频因其以用户所在的虚拟位置为球心，展示出整个球面场景而得名，其通过专业的 360° 全景摄像机拍摄，并利用相关技术以及软件处理后生成。更进一步而言，沉浸式的全景视频，即目前通过头戴式显示器（HMD）观看的全景视频，还需依靠立体成像技术方可具备良好的立体感和沉浸感，让用户感受到 360° 无死角的视频体验。

1.1 成像原理及技术

1.1.1 全景成像原理及技术

全景立体成像技术可为机器人导航、军事侦查、计算机视觉、虚拟现实等领域提供大视角场景的立体感知和重现功能，近年来发展快速，成为光电子学、计算机视觉和计算机图形学的研究热点。全景成像 (panoramic imaging, PI) 是指利用特殊的成像装置从一个视点获取水平方向一周 360°、垂直方向大到半球以上视场的多方向成像；而全景立体成像技术则是基于全景成像技术，利用双目视觉原理获取全景立体信息，该立体图像对的视差反映了场景的深度信息。

本节将就目前广泛应用的几种全景立体成像方法进行原理技术及应用的介绍。

全景立体成像特性

全景成像是在一个视点对周围所有方向的场景成像，而立体成像需要两个或多个视点，在垂直于两视点连线的方向上获得的场景具有立体感，但在两视点连线方向上不会具有立体感。

全景成像基本方法

全景成像系统一般由全景视觉光学成像系统、图像传感器、图像处理系统和输出图像显示器等部分组成。目前，全景成像的系统或方法主要有以下几种：

(1) 图像拼接成像

该方法使相机围绕其光心的垂直轴线旋转一周，进而对水平一周多个不同方向的场景成像，再将这些不同方向的场景图像进行拼接，获得一幅全景图像。因而，其中的硬件部分必须有精确的转动机构。图 1.1 所示的是应用该方法的旋转式云台相机。



图 1.1 云台相机

旋转式全景相机问世较早，1844 年德国人已经试验成功全景摇头转机。1920 年，Stromberg 等人发明了旋转式 360° 全景相机。1922 年，Richards 设计可以调整镜头和转轴的位置、底片放置方式的方案，研制出高质量旋转式全景相机。旋转式全景相机在中国使用的历史也比较长，主要是照相馆用其拍摄团体照，我国的上海、天津等地的照相机厂都曾生产过，至今仍有新的品种，但以胶片相机为主。目前国外已有彩色线阵 CCD 旋转式数码相机，在进口的同类相机中，瑞士生产的“环摄”转机最为著名，由瑞士 Seitz 公司生产的一款像素过亿的全景数码相机，像素可达 4 亿以上。这款 360 度全景相机每次拍摄大概要花费 2s 的时间，采用 80mm 的镜头时像素可达 4.7 亿。此相机拍摄相片的分辨率可达 7500x21250 像素。2004 年 1 月美国“勇气”号火星探测器登陆火星，上面安装有旋转式全景视觉系统。美国宇航局公布的“勇气”号火星探测器拍摄的火星三维全景黑白照及火星表面高分辨率全景彩照，都是由旋转式全景视觉系统获得。

该方法虽然成像分辨力高，但成像速度较慢，拼接算法复杂，不满足单一视点约束，一般只能拼接出圆柱面全景图像，而最大的不足是不能实时动态全景成像。

另一种基于图像拼接的方法是利用面向不同方向的多个相机同步拍摄多幅图像，进而将同步采集的多幅图像进行融合拼接得到全景图像。为获取水平方向 360° 和垂直方向一定角度的场景信息，根据不同方位镜头视场的大小确定所需相机个数。如要求该成像设备满足单一视点要求，则各台相机的光学中心必须重合。目前，在全景图像应用领域比较常见的就是这种多相机拼接式全景视觉系统，这类系统在军民领域均已得到了广泛的应用。

由此延伸的关于全景图像拼接方面的具体技术还将在后续章节中详细介绍。

(2) 鱼眼镜头成像

鱼眼镜头（图 1.2）是一种极端的超广角物镜，其焦距非常短（小于 16mm），能获取接近或大于 180° 的全景视场。鱼眼镜头由于获得大视场角，存在较大的桶形畸变，除了图像中心的景物保持不变，其它本应水平和垂直的景物沿各个方向从中心向外辐射，发生较大的畸变。



图 1.2 鱼眼镜头

鱼眼镜头的发展经历了从水下鱼眼的简单模仿到目前 270° 超大视场鱼眼镜头的过程。1919 年，Wood 在一个装满水的容器上盖一块玻璃板，构成鱼眼相机来实现超广角摄影，这套装置是对水下鱼眼最简单的模仿。1922 年 Bood 改进了 Wood 的装置，用半球玻璃透镜取代装满水的容器构成全景成像装置。1923 年，Hill 改进 Bood 的设计，在半球透镜前面引入一个光焦度绝对值很大的负弯月形透镜，改进后的镜头能够拍到较好的半球空域的云层照片，所以这种结构也叫做希尔天空镜头，之后的鱼眼镜头设计都沿用这种设计思想。后来一些学者继续对鱼眼镜头做了改进，提升了鱼眼镜头的成像质量。20 世纪 60 年代，光学自动设计技术的应用，使鱼眼镜头的发展更加迅速，出现了很多像质更加优良的光学结构。1964

年, Miyamoto 设计的镜头, $2\omega = 180^\circ$, $f = 16\text{mm}$, 不仅使系统像差得到了较好的校正, 并且像面照度的均匀性有了明显的改善。有些鱼眼镜头视场角能达到 220° , 有的甚至能达到 270° 。鱼眼镜头在摄影、医疗和安全监测等领域都有一定的应用。

高质量的鱼眼镜头通常需要采用 10 片以上的高质量光学材料结构, 因此还具有系统复杂、造价成本昂贵等缺点。

(3) 反射式成像

反射式全景视觉系统是由一台摄像机和一个反射曲面构成, 它是把反射曲面和传统的透射成像透镜组合在一起, 实现大视场成像的一种光学结构, 全景成像投影原理与鱼眼镜头全景成像的投影原理类似。摄像机垂直放置, 光轴垂直于水平面, 在摄像机上面放置一个曲面反射镜, 反射曲面将与光轴夹角很大的入射光线反射成为夹角较小的光线, 然后经摄像机成像, 从而增大视场。现有的应用于反射式全景视觉系统的反射面有抛物面、椭球面、双曲面等。根据反射面不同, 反射式全景视觉系统又可以分为抛物面反射式全景视觉系统、椭球面反射式全景视觉系统、双曲面反射式全景视觉系统等。

自从 1970 年美国宾夕法尼亚大学 W. Rees 设计了一套双曲反射镜面全景成像系统并成功应用于塔楼士兵监测周围目标后, 各国科研人员对反射式全景视觉系统进一步拓展。Yamazawa 等人利用双曲面反射镜与一个透镜相结合构成双曲面反射式全景成像系统。Chahl 等人在前人研究基础之上, 具体给出了双曲面反射镜的曲面公式。A. Gardel 等人则开展了反射式全景嵌入式系统的研究。美国 DARPA 将反射式全景视觉系统用于 Urbie 战术侦察机器人, 主要用于城市地形战术侦察。

国内反射式全景视觉系统的研究工作起步于 20 世纪 90 年代中期, 四川大学曾吉勇博士和苏显渝教授对抛物面反射镜、双曲面反射镜等全景视觉系统的原理、结构、成像特点和展开效果做了比较系统的研究。哈尔滨工程大学朱齐丹教授开展了高分辨率嵌入式反射全景视频处理系统的研究, 全景图像的分辨率为 2048×2048 。北京理工大学李科杰、张振海、王健等人开展了高分辨率高灵敏度抛物面反射式全景视觉系统研究, 作用距离约 1.5km, 有效全景图像的分辨率为 2750×2200 , 605 万像素, 帧率接近实时 25 帧, 最低照度可达 0.00031x 。

(4) 折反射式成像

折反射成像利用镜头前的反射镜扩大相机的视场从而获取实时全方位图像。这种成像方法, 能够实现实时大视场成像, 结构和几何计算简单, 立体图像对应点匹配容易等优点, 且成本较低。

折反射全景成像基本原理

折反射全景成像是利用曲面反射镜把水平方向 360° 范围内物体的光线反射到成像传感器, 从而一次性拍摄获得远大于普通相机视场范围的景象。与图像拼接和鱼眼镜头技术相比, 折反射全景成像技术能一次性实时获取 360° 图像, 具有大于半球空间的视场、成像装置设计简单、成本低等优点。折反射全景成像系统主要由三部分组成: 感光元件 (CCD 或 CMOS 器件)、成像透镜 (如常规成像透镜或远心透镜) 和反射镜, 如图 1.3 所示。由折反射全景成像系统获取的原始图像 (即球形图) 称为折反射全景图像, 如图 1.4 的上图所示。但由于原始图像存在严重变形, 不适合直接观察, 需要把折反射全景图像变换为符合人眼视觉习惯的圆柱面投影图像, 如图 1.4 的下图所示。根据是否满足单视点约束, 折反射全景成像系统分为单视点成像系统和非单视点成像系统。单视点成像的优势在于能将全景图变换成无畸变的平面透视图, 且能保证成像系统使用一台相机一次捕获到所需要的信息, 观察视角更大, 一般的图像处理方法可以用来进行图像分析和处理。

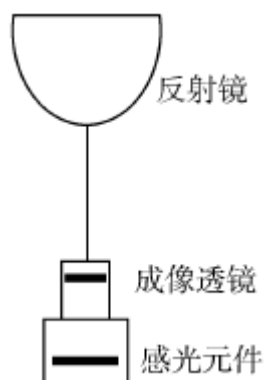


图 1.3 折反射全景成像系统组成

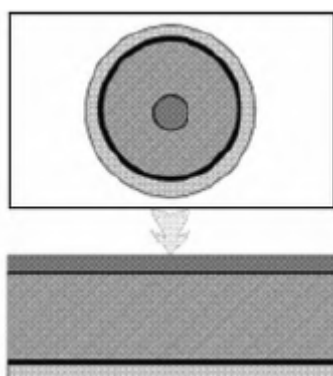


图 1.4 全景图像展开为柱面图像示意图

折反射全景成像反射镜面形式

折反射全景成像系统需要一个经过特殊加工的表面光滑的反射镜,将周围空间的光线汇集到反射镜上(分为凸面反射和凹面反射)。反射镜根据镜面线型的不同,又分为球面、双曲面、抛物面、圆锥面、椭球面等类型,如图 1.5 所示。

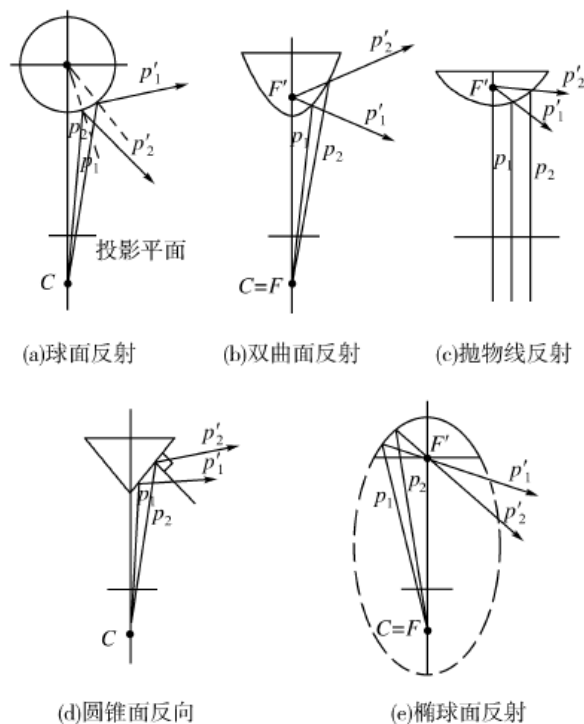


图 1.5 反射镜光路示意图

折反射成像映射原理

单视点全景成像系统,其所有入射光线的延长线都相交于同一视点,根据光路跟踪原理,对成像点进行光路跟踪和投影变换,可以建立单视点全景图像到柱面全景图像像素点间的坐标映射关系。如图 1.6 所示。把图 1.6(a)中的虚拟圆柱面沿与 x 轴正方向相交的一条母线展开,并以其左下角顶点为原点建立二维直角坐标系,可得图 1.6(c)所示的柱面全景图像,图 1.6(b)表示全景图像。根据不同的反射镜面类型和参数,推算的像素坐标映射关系也不同。另外,成像系统焦距、位置以及分辨率对像素坐标映射关系也有影响。

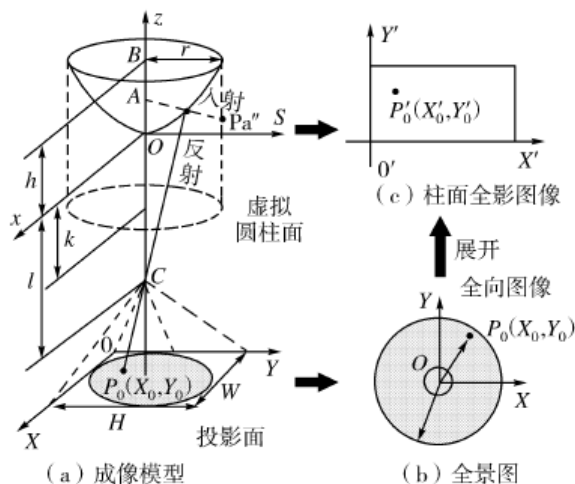


图 1.6 全景图转换为柱面展开图

折反射全景立体成像系统

利用 2 至 3 个折反射全景成像系统就可以实现全景立体成像。目前,使用两组折反射全景成像系统构成的立体成像系统装置结构主要有三种:(1)水平方向上并列放置两组独立的

全景成像装置，两组装置的轴线均垂直于水平方向，如图 1.7(a)所示。系统结构简单，但设备间会有一定相互遮挡，进而影响该遮挡区域真实场景的成像；(2)垂直方向上同轴上下放置两组独立的全景成像设备，如图 1.7(b)所示。虽然避免了第一种系统之间的相互遮挡，且立体图像对间存在良好的极线约束。但由于使用了两个成像设备，采集的立体图像对会引起曝光、颜色等方面的差别；(3)与第二种不同的是上下两组全景成像设备共用同一个相机，如图 1.7(c)所示。该系统要求相机拍摄到的一个反射面中要有另一反射面的图像，即在一幅图像中同时获取两个全景图像，避免了前两种系统使用两个相机两次成像导致的缺陷，但成像分辨率不高。

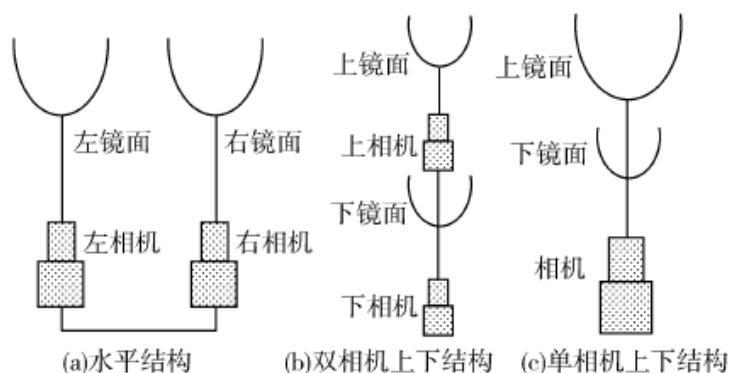


图 1.7 三种不同结构的折反射全景立体成像装置

使用三个折反射全景成像设备构成的全景立体成像系统装置结构如图 1.8 所示。三个折反射全景成像系统 A、B、C 分别放置在等边三角形的三个顶点上，每个全景成像系统水平方向的视场都是 360° ，均可等分为 I、II、III 三个区域。全景立体图像对取自 A、B、C 三个系统的不同视场区域。视场区域 A_I 、 B_I 、 C_I 和 A_{III} 、 B_{III} 、 C_{III} 分别组成了完整的左右眼全景图像，而 A_{II} 、 B_{II} 、 C_{II} 为相互遮挡的区域。该全景立体成像数据处理相对简单，能实现对动态景物的全景立体成像，但硬件结构较复杂，成本较高。

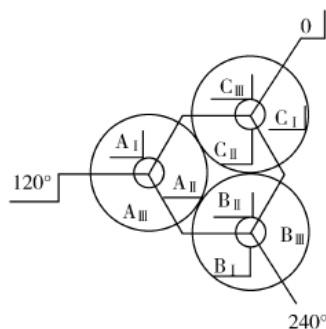


图 1.8 全景立体成像系统

Greguss 教授首先提出 PAL 全景透镜的结构，并进一步提出使用抛物面设计 PAL 的两个透射面和两个反射面，得到更精准的成像结果，并尝试设计全景环带透镜光学系统。进入 90 年代，Powell 对全景环带光学系统的结构开展深入研究，首次提出全球面 PAL 结构，与此同时，美国阿拉巴马大学 Gilbert 教授、Fair 教授与马奎特大学 Matthys 教授等人对 PAL 光学系统结构开展仿真计算，并对 PAL 光学系统的应用做了深入探讨，并详细讨论了物面测绘定位和环带像面展开的方法。索尼公司研制 360° 折反式全景相机分辨率约为 38 万像素，通过 DSP 全景图像算法程序，最终输出图像输出 7.5 帧 CCD 全景图像。这种新型的全景摄影装置主要用于监控领域。

国内开展折反式全景视觉系统研究的单位主要有浙江大学、北京理工大学等。浙江大学

开展了较长焦距全景环带成像光学系统研究,设计的小口径 PAL 结构的视场角为 50° 全景成像光学镜头。北京理工大学李科杰、张振海首次研制折反射式微光全景视觉系统,成像视场大小为水平方向 360° ,垂直方向仰角 37° ,俯角 18° ,最低照度可达 $0.00031x$,可实时输出 PAL 标准制式微光全景图像,通过研制 FPGA+DSP 嵌入式全景图像系统,将全景环形图像展开、畸变校正为 8 个方向全景图像,可用于地面侦察机器人微光条件下的全景图像信息获取。

小结

全景成像技术由于其视场范围大的优点,广泛应用于机器人导航、军事侦查、计算机视觉、虚拟现实等领域。本节介绍了几种主要的全景立体成像技术,包括旋转图像拼接、鱼眼镜头、反射和折反射成像四种方法及其各自优缺点。用于生成沉浸式全景视频的典型和新型全景相机/系统将在 1.4 节中进行详细介绍。

1.1.2 3D 成像原理及技术

现实世界是一个三维立体世界,随着社会的发展,传统的二维平面显示在某些方面已不能满足人类的需求,人们希望显示器或其他设备能真实地还原显示出空间的三维信息。因此,三维立体显示应运而生,并不断得到发展,一度成为显示领域的一个研究热点。

19 世纪 30 年代,科学家 Wheatstone 着手研究人的视觉,并于 1838 年发明了立体镜,拉开了人类对三维立体显示研究的帷幕。20 世纪中后期,随着计算机信息领域的发展以及平板显示器的出现,三维立体显示蓬勃发展。日本、韩国、欧美等国家从 20 世纪 80 年代开始了三维立体显示的基础研究,各国根据自身的情况,开发了各种技术和产品。我国三维立体显示虽起步较晚,基础薄弱,与国际水平存在差距,但也有不少高校、研究所和公司企业对立体显示开展研究并取得了一些成果。

三维立体显示目前被广泛应用,而对于需要立体感、沉浸感的全景视频而言,该技术同样关键。如果说黑白显示器是第一代产品,彩色显示器是第二代产品,那么立体显示器就是第三代产品。自 19 世纪 30 年代开始,经过接近两个世纪的发展,已开发出各种三维立体显示。本节按其基本工作原理是否为双目视差将三维立体显示分为两大类。所谓双目视差是指人两眼间有一定瞳距,大约有 6-7 厘米的间隔,在观看物体时左眼和右眼所接收到的视觉图像略有差异。基于双目视差原理的三维立体显示为观看者的左右眼提供同一场景的立体图像对,采用光学等手段让观看者的左右眼分别只看到对应的左右眼图像,这样便使观看者感知到立体图像。这类三维立体显示的技术相对成熟并有相应产品,但存在观看视疲劳等问题。非基于双目视差原理的各种三维立体显示的工作原理各不相同,如利用光学干涉衍射原理、人眼视觉暂留效应以及人眼视错原理等。这类三维立体显示不存在观看视疲劳,但技术相对而言尚未成熟。下面对这些三维立体显示技术作简要阐述。

基于双目视差原理的三维立体显示

早在 19 世纪摄影技术刚刚起步时,人们就用两台性能和参数完全相同的相机并列,模拟人的左右双眼,同时拍下两张有微小差异的相片,之后再透过平行视线法、交叉视线法或类似双筒望远镜的专属观看设备等,让人的左右双眼分别观看两张并列拍摄的相片,以重现视差,借以模拟出立体视觉。

目前,基于双目视差原理的三维立体显示技术主要有眼镜/头盔式立体显示和光栅式自由立体显示两类,前者技术和产品都较为成熟,后者已有一些产品,但其性能大多有待提高。

(1) 眼镜/头盔式三维立体显示

眼镜三维立体显示按其工作原理主要分为三类：

一是基于波长，如红绿、红蓝等互补色立体眼镜的三维立体显示，其利用人类能感知红蓝绿三原色的机理，使进入左右眼的光谱不同，从而形成双眼视差。以红蓝光眼镜为例，片源是由红、蓝光染色生成的具有双目视差的立体图像，由于相同颜色镜片会过滤掉相同颜色的图像。因此，通过红色镜片观察的眼睛，只能看到蓝色图像，同样通过蓝色镜片观察的眼睛，只能看到红色图像，由于红蓝光产生的图像本身具有差异，从而产生立体视觉。其主要特点是技术成熟、成本低，但不能显示彩色图像。



图 1.9 红蓝 3D 眼镜

二是基于时序立体眼镜的三维立体显示，其显示屏分时显示左右视差图，并通过同步信号发射器及同步信号接收器控制观看者所佩戴的液晶快门立体眼镜，使得当显示屏显示左(右)眼视差图像时左(右)眼镜片透光而右(左)眼镜片不透光，按照上述方法将两套画面以极快的速度切换，在人眼视觉暂留特性的作用下就合成了连续的画面。其主要特点是要求显示器的帧频为普通显示器的两倍，一般要达到 120 Hz。

三是基于偏振眼镜的三维立体显示，显示屏上左右眼视差图的光线为互相正交的线偏振光或左右旋圆偏振光，通过偏光滤镜或偏光片滤除特定角度偏振光以外的所有光，让 0 度的偏振光只进入右眼，90 度偏振光只进入左眼，即双眼分别看到立体图像对中不同的图像，因而观众佩戴上相对应的偏光眼镜就可以观看到立体效果。



图 1.10 偏振式 3D 眼镜

头盔式三维立体显示是在观看者双眼前各放置一个显示屏，观看者的左右眼只能分别观看到显示在对应屏上的左右视差图，同时遮挡住外部的光线和视野，从而提供给观看者一种沉浸于虚拟世界的沉浸感。显示器一般与电脑或手机的视频输出连接，分别提供给左右眼不同的图像。头盔显示器提供一个稳定的双眼视差三维影像，利用附带的头部跟踪器后可实时获取头部的位置和运动信息传递给生成图像的电脑或手机，这样就可以根据观察者的位置和方位信息提供不同图像画面。目前，全景视频的成像、显示、播放主要通过 HMD 以及相关设备完成。这种立体显示存在单用户性、显示屏分辨率低、头盔沉重及易给用户带来不适感等缺点。



图 1.11 用于观看沉浸式全景视频的 HMD

(2) 光栅式三维自由立体显示

光栅式自由立体显示器主要是由平板显示屏和光栅精密组合而成，左右眼视差图像按一定规律排列并显示在平板显示屏上，然后利用光栅的分光作用将左右眼视差图像的光线向不同方向传播，当观看者位于合适的观看区域时其左右眼分别观看到左右眼视差图像，经过大脑融合便可观看到有立体感的图像。根据采用的光栅类型可分为狭缝光栅式自由立体显示和柱透镜光栅式自由立体显示两类。

狭缝光栅式自由立体显示器又分为前置狭缝光栅和后置狭缝光栅两种，其结构与原理图分别如图 1.12(a) 和图 1.12(b) 所示。图 1.12(a) 中狭缝光栅置于平板显示屏与观看者之间，观看者左右眼透过狭缝光栅的透光部分只能看到对应的左右眼视差图像，由此产生立体视觉。图 1.12(b) 中狭缝光栅置于平板显示屏与背光源之间，用来将背光源调制成狭缝光源。当观看者位于合适的观看区域时，从左(右)眼处看显示屏上只有左(右)眼视差图像被狭缝光源照亮，那么左右眼就只能看到对应的左右眼视差图像，由此产生立体视觉。

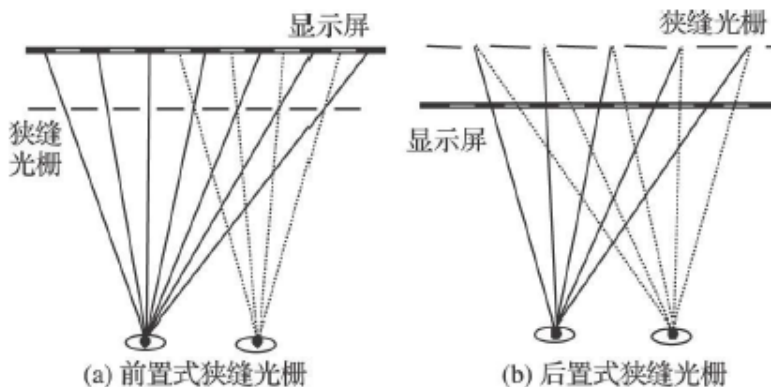


图 1.12 狭缝光栅式自由立体显示器的结构与原理

柱透镜光栅式自由立体显示器的结构与原理如图 1.13 所示，利用柱透镜阵列对光线的折射作用，将左右眼视差图像分别提供给观看者的左右眼，经过大脑融合后产生具有纵深感的立体图像。

狭缝光栅式自由立体显示由于狭缝不透光部分对光线的遮挡，从而导致立体图像亮度相对于平面图像损失严重；而柱透镜光栅式自由立体显示由于采用透明的柱透镜，除了透镜对光线的吸收外立体图像亮度基本没有损失。这两种光栅式立体显示由于在平板显示器上同时显示两幅或多幅视差图像，从而导致立体图像分辨率相对于平面图像有所降低。光栅式自由立体显示由于其结构简单、易于实现、无需佩戴辅助观看装置及立体显示效果良好等优点得到立体显示研究人员的青睐。然而要使光栅式自由立体显示得到广泛应用，还需要研究人员不断提高其性能，改善其立体显示效果。

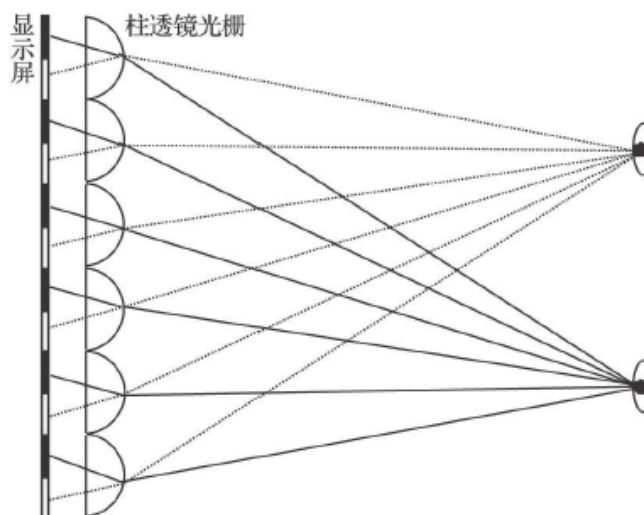


图 1.13 柱透镜光栅式自由立体显示器的结构与原理

(3) 自动分光立体显示

近年来，不需要戴眼镜或头盔的自动分光立体显示器已经从实验室进入市场，大多数自动分光立体显示器是基于平板显示器的系统，通过结合水平的双眼视差和运动视差给出了不戴眼镜的 3D 立体效果。人们在真实的世界里观察主要有两种视觉感受，一是双眼会看到不同画面的双眼视差，二是头部处于不同位置或运动都会看到不同的画面，相当于沉浸在一组无数张画面组成的场景。自动分光立体显示技术便利用到上述的视觉特性，通过将原视景划分为几个视景组，使得人双眼落在不同的视景组中，左右便会看到不同画面，从而产生双眼视差。同样，处于不同位置或运动，人眼视线也同时落在不同视景组里，便能实现水平运动视差。该技术特点是不需要头部跟踪以及相关的复杂技术，多个观察者可同时观察区域中自由地移动双眼，获取多视角的 3D 效果。然而，建立自动分光系统比较困难，同时产生所有视图可能使观察者看见错误的图像，且偏离理想的距离越远，这种可能性更大。

非基于双目视差原理的三维立体显示

非基于双目视差原理的三维立体显示主要有全息立体显示、集成成像立体显示以及体显示等，它们的技术和产品尚不成熟，需要开展更深入的研究工作。

(1) 全息立体显示

全息技术是利用干涉原理将物体发出的特定光波以干涉条纹的形式记录下来，形成“全息图”，全息图中包含了物光波前的振幅及相位信息。当用相干光源照射全息图时，基于衍射原理重现原始物光波，从而形成原物体逼真三维图像。全息立体显示是一种真三维立体显示技术，观看全息立体图像时具有观看真实物体一样的立体感。全息图的每一部分都记录了物体各点的光信息，故即使全息图有所损坏也照样能再现原物体的整个图像。通过多次曝光可在同一张底片上记录多个不同图像且互不干扰地分别显示出来。

近年来，随着计算机技术的发展和高分辨率电荷耦合成像器件 (Charge Couple Device, CCD) 的出现，数字全息技术得到迅速发展。与传统全息不同的是，数字全息用 CCD 代替普通全息记录材料记录全息图，用计算机模拟取代光学衍射来实现物体再现，实现了全息图记录、存储、处理和再现全过程的数字化。数字全息技术和显示技术的结合使全息技术的应用前景更加广阔，为数字全息在真 3D 显示中的应用奠定了良好的基础。目前，数字全息技术主要有三种：基于数字合成全息技术的三维显示技术、基于空间光调制器的数字全息三维显

示技术、基于集成技术的数字全息三维显示技术。

(2) 集成成像立体显示

集成成像技术也由记录和再现两个基本过程组成，与全息技术不同的是其记录再现过程并不需要相干光的参与，而是利用二维微透镜阵列来实现。如图 1.14(a) 所示，第一步将记录介质置于微透镜板的焦平面上，每个微透镜单元从不同的方向记录物体空间的部分信息，将每个微透镜单元所记录的一幅幅小图像称为“子图像”，空间物体的视差信息就被这一幅幅“子图像”记录下来。第二步将记录介质置于具有相同参数的微透镜阵列的焦平面上，用散射光照射，由光线可逆原理就可在微透镜板另一侧观看到再现物体空间场景。如图 1.14(b) 所示。

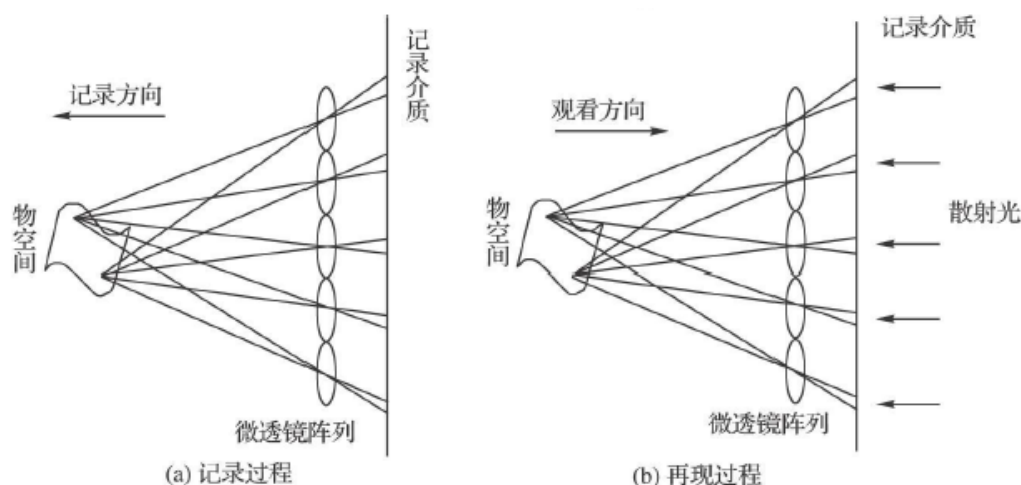


图 1.14 集成成像技术的记录和再现过程

利用集成成像技术实现三维立体显示可供多个观看者同时观看且无需配戴特殊眼镜，观看三维图像时不存在眼睛会聚与调节不匹配的问题，再现的三维场景具有全真色彩以及连续视差。但集成成像技术也具有其自身的不足之处需要改善：1) 记录和再现两过程中会产生空间反转，通过两次记录可解决空间反转问题，但图像质量会有所下降，采用负透镜阵列等技术可较好地解决空间反转问题；2) 立体观看视角比较窄，可通过透镜开关及非球面透镜等方法来拓宽立体视角；3) 可清晰记录的物体景深比较小，一般为几个厘米，可利用双图像平面等方法来增强景深；4) 分辨率低，可采用逆光线获取等方法来提高分辨率。此外，子图像间的相互串扰以及透镜单元本身的像差等因素也会影响到集成成像质量。随着对集成成像理论研究的不断深入以及对成像质量的不断改善，集成成像技术将成为立体显示领域的重要研究方向。

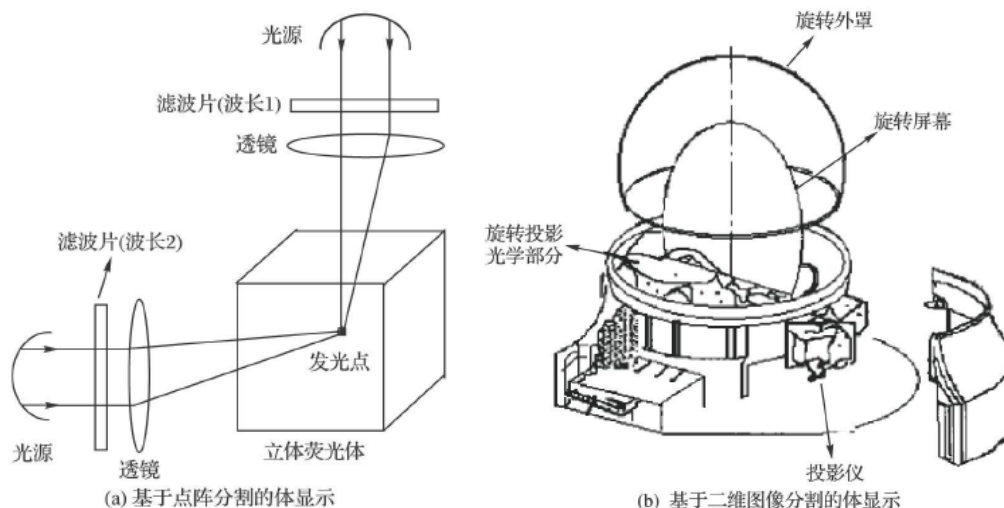


图 1.15 体显示的工作原理

(3) 体显示

近年来，随着计算机数据处理能力的迅速提高和数据存储技术的发展，采用嵌入式系统的 3D 立体显示器引起人们的广泛关注。这种显示器利用屏幕的旋转和光投影等技术，将原来的二维图像切片合成富有真实立体感的三维图像，和真实物体的视觉效果比较接近。可以让多人从多角度观察模型(有的系统支持人机交互功能)。其娱乐性、舒适度及自然性也较好。

体显示通常是将三维物体分割为点阵或一系列二维图像，再依次扫描，利用人眼的视觉暂留效应形成立体图像。图 1.15(a)中是把三维物体分割为点阵再依次扫描，图中所示立方体是添加了发光物质的透明荧光体，两束不可见波长的光聚焦到同一点进行激发，从而发出可见光，对立方体中每点依次扫描即可形成立体图像。图 1.15(b)中是把三维物体分割为二维图像再依次扫描，以半圆形显示屏作为投影面，将它高速旋转，在空间形成一个半球形成像区域，在旋转的过程中将半圆形显示屏像素有规律的点亮，由于人眼视觉暂留效应从而观看到空间连续的三维图像。体显示可供多个观看者同时从不同角度观看到同一显示图像的不同侧面，且兼顾了人眼的调节和会聚特性，不会引起视觉疲劳。

小结

本节按基本工作原理是否为双目视差将三维立体显示分为两大类，并对各种三维立体显示器的结构、基本原理以及优缺点进行了简要分析。和一些研究方向不同，三维立体显示具有百家争鸣、百花齐放的特点，各种三维立体显示各具特点，各有各的使用场合；除了目前热门的全景视频、增强现实研究领域外，3D 成像的概念也已应用到了立体电影、音像制品、手机、立体电视、户外媒体等产品与服务中。与其他新兴研究方向一样，立体显示技术还不成熟，还有很多未知的认识需要探讨，更高性能的器件和系统等待研制并产业化。我们也可以期待，在未来，人们可以脱去沉重的 HMD 轻松观看全景视频。

1.2 全景图像拼接技术

近年来，随着计算机技术的快速发展，图像融合技术发展越来越广泛和深入，对具有较大视域全景图像的需求也越来越迫切。全景图像拼接作为新兴技术，短短几年得到了快速发展，受到研究者越来越多的关注。目前全景图像已经成为计算机仿真、计算机视觉模拟、图

像处理和计算机特效以及虚拟现实研究中的热点和关键技术，在地质勘测、军事侦查、医学微创手术、航空航天以及视频会议等多个领域发挥着重要作用。

在之前章节中提到，目前全景图像/视频主要通过多相机拼接式全景视觉系统拍摄，由于相机多角度的特性，这类系统在产生原始图像后，需要通过全景图像/视频拼接技术进行后处理才能生成完整、统一的全景内容。

所谓图像拼接是指把多个单一图像融合成一幅图像。具体地说，全景拼接是将使用多个摄像机对同一个场景在不同角度拍摄得到的多个图像进行校正、去噪、匹配、融合，最终构建成一个质量高、清晰、边缘平滑、分辨率高的图像。图像拼接主要有以下几个步骤：相机标定、图像坐标变换、图像畸变校正、图像配准与融合。其中，图像配准与融合是图像拼接成功的关键。

全景拼接实现过程

(1) 相机标定

全景视觉系统中安装设计，以及摄像机之间的差异，会造成图像/视频之间有缩放（镜头焦距不一致造成）、倾斜（垂直旋转）、方位角（水平旋转）的差异，这类物理差异需要预先标定和校准，才能得到一致性好的图像，便于后续图像拼接。相机标定主要是求解相机的内参数矩阵，用于后续图像处理。该矩阵主要与相机的运动方式有关，常用求解方式是张正友标定法。该方法精度高，又避免了传统标定法对标定器材要求高、操作过程繁琐的约束。

相机的运动方式与成像结果之间的关系见下图：

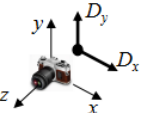
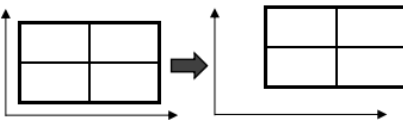
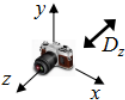
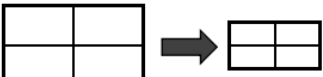
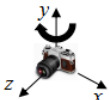
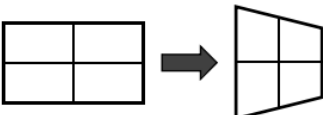

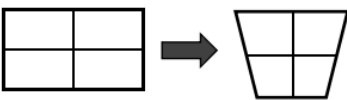
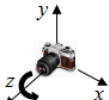
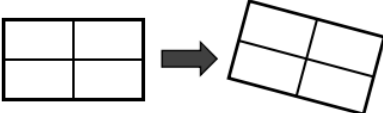
名称	相机运动示意	图像变化结果	图像变换
平移			平移变换
变焦			放缩变换
水平旋转			投影变换
垂直旋转			投影变换
绕轴旋转			旋转变换

图 1.16 相机的运动方式与成像结果之间的关系

(2) 图像坐标变换

在实际应用中，全景图像的获得往往需要摄像机以不同的位置排列和不同的倾角拍摄。例如由于机载或车载特性，相机的排列方式不尽相同，不能保证相机在同一面上，如柱面投影不一定在同一个柱面上，平面投影不一定在同一平面上；另外为了避免出现盲区，相机拍

摄的时候往往会向下倾斜一定角度。这种情况比较常见，而且容易被忽略，直接投影再拼接效果较差。因而有必要在所有图像投影到某个柱面（或平面）之前，需要根据相机的位置信息和角度信息来获得坐标变换后的图像。

理论上只要满足静止三维图像或者平面场景的两个条件中的任何一个，两幅图像的对应关系就可以用投影变换矩阵表示，换句话说只要满足这其中任何一个条件，一个相机拍摄的图像可以通过坐标变换表示为另一个虚拟相机拍摄的图像。

为了确定图像序列的空间变换关系，需要确定图像对应关系的模型。平移变换模型、刚性变换模型、仿射变换模型和投影变换模型是目前常用的几何图像变换模型，图 1.17 即为几种几何变换模型示意图。

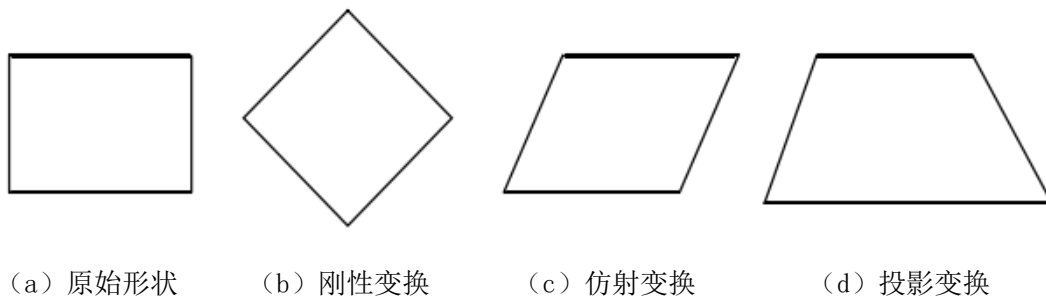


图 1.17 图像变换模型示意图

一般情况下，8 参数的透视投影变换最适合描述图像之间的坐标关系，其中 8 参数的矩阵为 $[m_0, m_1, m_2; m_3, m_4, m_5; m_6, m_7, 1]$ ，相应几何关系使用矩阵的形式描述出来：

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = M \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1.1)$$

各参数对应的相机运动表示如下：

表 1.1 透视投影变换参数与相机运动的关系

参数	对应摄像机的运动及成像结果
m_2	x 方向位移
m_5	y 方向位移
m_0, m_1, m_3, m_4	缩放，旋转，剪切
m_6, m_7	梯形失真（x 方向和 y 方向形变），线性调频

如图 1.18 显示的是相机向下倾斜一定角度拍摄图像的情况，这个角度与 m_6 和 m_7 具有对应关系，只需要对 8 参数矩阵求逆后进行坐标变换，即可获得校正图像，。



(a)



(b)



(c)

图 1.18 (a) 原始图像；(b) x 方向形变效果；(c) 倾斜校正后效果

(3) 图像畸变矫正

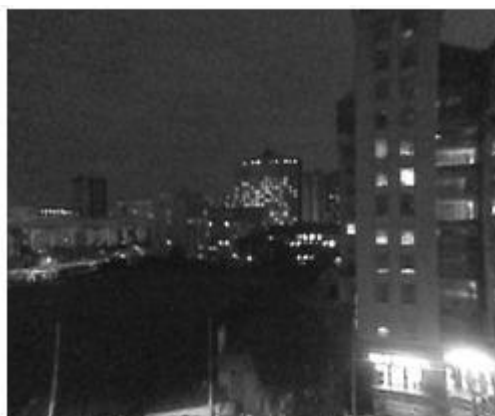
由于制造、安装、工艺等原因，镜头存在着各种畸变，为了提高摄像机拼接的精度，在进行图像拼接的时候必须考虑成像镜头的畸变。一般畸变分为内部畸变和外部畸变，内部畸变是由于摄像本身的构造为起因的畸变，外部畸变为投影方式的几何因素起因的畸变。镜头畸变属于内部畸变，由镜头产生的畸变一般可分为径向畸变和切向畸变两类。径向畸变就是集合光学中的畸变像差，主要是由于镜头的径向曲率不同而造成的，有桶形畸变和枕形畸变两种。切向畸变通常被认为是由于镜头透镜组的光学中心不共线引起的，包括有各种生成误差和装配误差等。一般认为，光学系统成像过程中，径向畸变是导致图像畸变的主要因素。径向畸变导致图像内直线成弯曲的像，且越靠近边缘这种效果越明显。根据径向畸变产生的机理，对视频图像进行校正。效果如图 1.19 (b) 所示，经过校正的图像，其有效像素区域缩小，一般可通过电子放大的方式进行校正，如图 1.19 (c) 所示。



(a)



(b)



(c)

图 1.19 (a) 原始采集图像；(b) 经过径向失真校正的图像；(c) 经过放大的图像

(4) 图像配准与融合

全景拼接过程的最后一步，也是最为关键的一步是图像配准与融合。图像配准与融合在全景图生成、360° 全景相机以及 VR 全景领域有非常多的应用，常用的图像缝合工具有 Microsoft 的 ICE、PTGui、开源软件 Hugin 等，基于视频的拼接可以参考 VideoStitch、StitchHD 以及 stitching_with_cuda。

一般的图像配准与融合算法步骤可以描述为：

- 1) 定义映射模型，常用的包括：球面、柱面、平面，其中球面映射应用最为广泛；
- 2) 根据输入图像，提取特征点，对特征进行匹配，得到输入图像之间的映射关系 T ；
- 3) 根据映射关系 T 进行图像的 Warp 变换，对齐图像；
- 4) 包括利用颜色调整来消除图像间的色差，和采用图像融合来消除拼缝。

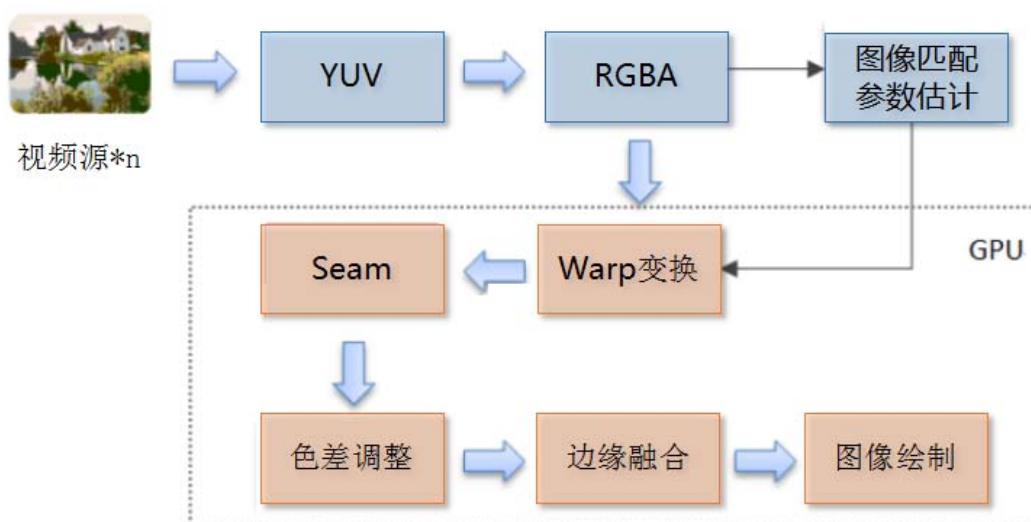


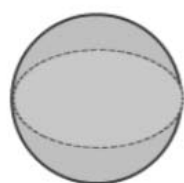
图 1.20 图像融合算法流程

图像投影变换

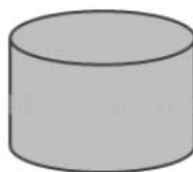
由于每幅图像是相机在不同角度下拍摄得到的，所以它们并不在同一投影平面上，如果对重叠的图像直接进行无缝拼接，会破坏实际景物的视觉一致性。所以需要先对图像进行投影变换，再进行拼接。一般有平面投影、柱面投影、立方体投影和球面投影等。

平面投影就是以序列图像中一幅图像的坐标系为基准，将其图像都投影变换到这个基准坐标系中，使相邻图像的重叠区对齐，称由此形成的拼接为平面投影拼接；柱面投影是指采集到的图像数据重投影到一个以相机焦距为半径的柱面，在柱面上进行全景区的投影拼接；球面投影是模拟人眼观察的特性，将图像信息通过透视变换投影到眼球部分，构造出一个观察的球面；立方体投影是为了解决球面映射中存在的图像不宜存储的缺点，而发展出来的一种投影拼接方式，它适用于计算机生成图像，对实景拍摄的图像则比较困难。

投影模型可以看作是用于图像映射的载体，相当于二维图像映射到三维空间的一种变换，如下图所示：



(a) 球面投影模型



(b) 柱面投影模型



(c) 平面投影模型

图 1.21 不同类型的投影模型

其对应拼接效果如下图所示：



图 1.22 不同投影模型的拼接效果

可以看到，不同投影模型对应的拼接效果是有所区别的，因而选择合适的投影模型非常重要。映射模型应与图像采集场景以及应用方式相匹配，一般对于水平拼接而言，采用柱面投影模型重现的效果最佳，而对于 360° 全景，球面映射或者立方体（多面体）映射的效果更好。由于上图中采用的是三张平行拍摄的图片，垂直方向张角较小，因此球面模型与柱面模型的拼接效果差异并不大。

特征点提取与匹配

由于特征点的方法较容易处理图像之间旋转、仿射、透视等变换关系，因而经常被使用，特征点包括图像的角点以及相对于其领域表现出某种奇异性的兴趣点。Harris 等人提出了一种角点检测算法，该算法是公认的比较好的角点检测算法，具有刚性变换不变性，并在一定程度上具有仿射变换不变性，但该算法不具有缩放变换不变性。针对这样的缺点，Lowe 提出了具有缩放不变性的 SIFT 特征点。

SIFT 特征检测主要包括以下 4 个基本步骤：

- 1) 尺度空间极值检测：搜索所有尺度上的图像位置。通过高斯微分函数来识别潜在的对于尺度和旋转不变的兴趣点；
- 2) 关键点定位：在每个候选的位置上，通过一个拟合精细的模型来确定位置和尺度。关键点的选择依据于它们的稳定程度；
- 3) 方向确定：基于图像局部的梯度方向，分配给每个关键点位置一个或多个方向。所有后面的对图像数据的操作都相对于关键点的方向、尺度和位置进行变换，从而提供对于这

些变换的不变性；

4) 关键点描述：在每个关键点周围的邻域内，在选定的尺度上测量图像局部的梯度。这些梯度被转换成一种表示，这种表示允许比较大的局部形状的变形和光照变化。



(a) 极值点初步检测

(b) 精确定位

图 1.23 SIFT 特征检测

特征匹配则用来计算图像之间的映射关系（采用 RANSAC 或者概率模型），得到每个匹配图像对之间的单应矩阵，结合上一步的映射模型，可以得到最终的图像变换序列 $T_1 T_2 T_3 \dots$ 。

不管使用何种图像检测算子，都需要在图像序列中找到有效的特征匹配点。图像的特征点寻找直接影响图像拼接的精度和效率。对于图像序列，如果特征点个数大于等于 4 个，则很容易自动标定图像匹配点；如果特征点很少，图像拼接往往不能取得较为理想的效果。

在计算映射变换之后，实际上相当于就得到了图像之间的全景变换关系，或者叫做相机变换参数（参照 OpenCV 里的 `detail::CameraParams`），对于视频拼接来讲，这个变换参数通常是不变的。

此外，也可以利用手动的方式进行调整，只要确保图像之间对齐即可，因此特征的提取与匹配并不是必选项。

图像配准

图像拼接中的一个关键步骤是配准，即将两幅待拼接图像对齐，找到两幅图像相对的位置关系。配准的目的是基于匹配的特征点，根据几何运动模型将图像转换到同一个坐标系中。在多幅图像配准的过程中，采用的几何运动模型主要有：平移模型、相似性模型、仿射模型和透视模型。

图像的平移模型是指图像仅在两维空间发生了方向和方向的位移，如果摄像机仅仅发生了平移运动，则可以采用平移模型。图像的相似性模型是指摄像机本身除了平移运动外还可能发生旋转运动，同时，当场景存在缩放时，还可以利用缩放因子对缩放运动进行描述，因此，当图像可能发生平移、旋转、缩放运动时，可以采用相似性模型。图像的仿射模型是一个 6 参数的变换模型，即具有平行线变换成平行线，有限点映射到有限点的一般特性，具体表现可以是各个方向尺度变换系数一致的均匀尺度变换或变换系数不一致的非均匀尺度变换及剪切变换等，可以描述平移运动、旋转运动以及小范围的缩放和变形。图像的透视模型是具有 8 个参数的变换模型，功能是完成从二维坐标到二维坐标之间的线性变换，且保持二维图形的“平直性”和“平行性”。其可以完美地表述各种变换，包括平移，缩放，翻转，旋转和剪切，是一种最为精确变换模型。

采用矩阵的形式可以描述为：

$$p' = H * p \quad (1.2)$$

$$\text{其中 } H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix}$$

矩阵 H 被称为投影变换矩阵。对上式进行分解化简，即是通常所谓的单应性变换：

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + 1} \quad (1.3)$$

$$y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + 1} \quad (1.4)$$

图像配准变换是一项耗时的操作，相当于对里面每一个像素点进行一次变换，像素之间的操作相互独立，因此操作通常放在 GPU 上来并行处理，对于 CUDA 来讲，透视变换相当于将输入图像的索引坐标值映射到纹理坐标。

亮度与颜色均衡处理

相机和光照强度的差异会造成一幅图像内部，以及图像之间亮度的不均匀，拼接后的图像会出现明暗交替的现象，极大地影响融合效果。因而在融合前，还需进行亮度与颜色均衡处理操作，通常的处理方式是通过相机的光照模型，校正一幅图像内部的光照不均匀性，然后通过相邻两幅图像重叠区域之间的关系，建立相邻两幅图像之间直方图映射表，通过映射表对两幅图像做整体的映射变换，最终达到整体的亮度和颜色的一致性。

具体的亮度与颜色均衡处理方法有很多，如自动白平衡方法，也可以手动调整色温，其基本思路都是通过统计每幅图的亮度或颜色区间分布，将不同的亮度、颜色调整到与参考颜色一致的空间内。

比较常用的是 Reinhard 方法，将图像 I 转换到 lab 空间（降低三原色之间的相关性），通过图像的统计分析，利用目标图像 I' 的均值及标准差进行线性调整，公式描述为：

$$\begin{cases} L = (l - \bar{l}) * \frac{\sigma_{l'}}{\sigma_l} + \bar{l}' \\ A = (a - \bar{a}) * \frac{\sigma_{a'}}{\sigma_a} + \bar{a}' \\ B = (b - \bar{b}) * \frac{\sigma_{b'}}{\sigma_b} + \bar{b}' \end{cases} \quad (1.5)$$

线性变换（方差作为斜率）能够保证源图像能够与目标图像在 lab 颜色空间具有近似的均值和方差。通常我们可以选定一副图像作为调整基准，当然也可以计算需要变换的所有图像的均值作为一个目标值。

全景图像融合与生成

图像融合则是将配准后的图像合成为一张大的拼接图像。待融合图像已配准好且像素位宽一致，综合和提取两个或多个多源图像信息。两幅（多幅）已配准好且像素位宽一致的待融合源图像，如果配准不好且像素位宽不一致，其融合效果也不好。

高效的图像融合方法可以根据需要综合处理多源通道的信息，从而有效地提高了图像信息的利用率，系统对目标探测识别的可靠性及系统的自动化程度。其目的是将不同相机以及传感器所提供的信息加以综合，消除多相机信息之间可能存在的冗余和矛盾，以增强影像中信息透明度，改善解译的精度、可靠性以及使用率，以形成对目标的清晰、完整、准确的信息描述。这诸多方面的优点使得图像融合在医学、遥感、计算机视觉、气象预报及军事目标识别等方面的应用潜力得到充分认识，尤其在计算机视觉方面。

一般情况下，图像融合由低到高分三个层次：数据级融合、特征级融合、决策级融合。数据级融合也称像素级融合，是指直接对传感器采集得到的数据进行处理而获得融合图像的过程，它是高层次图像融合的基础，也是目前图像融合研究的重点之一。这种融合的优点是保持尽可能多的现场原始数据，提供其它融合层次所不能提供的细微信息。

像素级融合中有空间域算法和变换域算法，空间域算法中又有多种融合规则与方法，如逻辑滤波法，灰度加权平均法，对比调制法等；变换域中又有金字塔分解融合法，小波变换法。其中的小波变换是当前最重要，最常用的方法。

其中，变换域中的拉普拉斯金字塔法可以理解为通过对相邻两层的高斯金字塔进行差分，将原图分解成不同尺度（频率）的子图，对每一个子图（对应不同频带）进行加权平均，得到每一层的融合结果，最后进行金字塔的反向重建，得到最终融合效果。

在特征级融合中，保证不同图像包含信息的特征，如红外光对于对象热量的表征，可见光对于对象亮度的表征等等。

决策级融合主要在于主观的要求，同样也有一些规则，如贝叶斯法，D-S 证据法和表决法等。

融合算法常结合图像的平均值、熵值、标准偏差、平均梯度；平均梯度反映了图像中的微小细节反差与纹理变化特征，同时也反映了图像的清晰度。

关于图像融合，目前已有现成的开源代码，这里不再过多描述。融合的关键在于选择用于融合的子图像区域部分，要注意图像过大导致的效率问题，也要避免图像较小带来的信息缺失现象。

OpenCV 的代码是图像/视频融合比较好的入门材料，可以作为参考，OpenCV 的流程框架图如下：

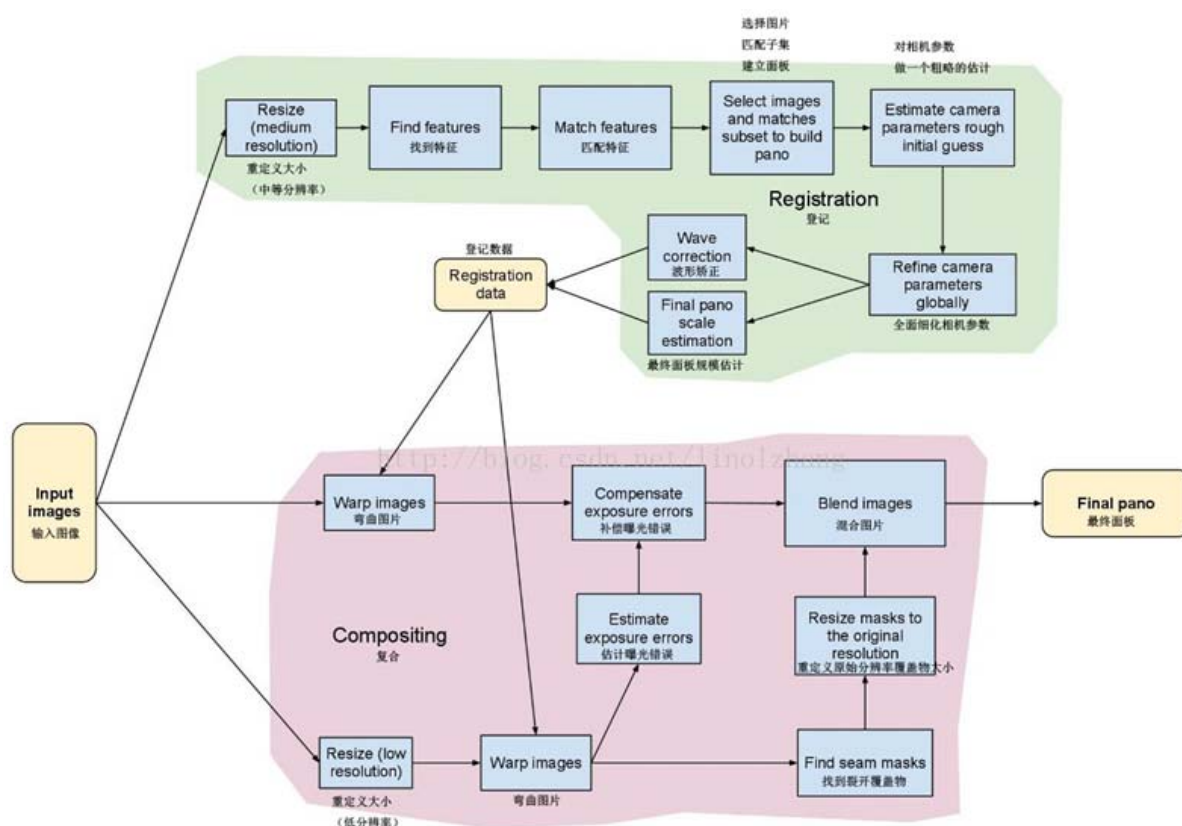


图 1.24 OpenCV 框架

1.3 开源软件

实际上，全景视频不仅可以用现实中的 3D 摄像机进行拍摄，也可以直接用三维软件进行渲染做出全景 CG (Computer Graphics)。对于最终可观看的全景视频的实现，可以通过将球型摄像机拍摄的或 360° 多角度拍摄的视频直接映射到一个球面上，也可以将拍摄的视频通过 EPR 映射将球形画面映射到矩形画面中，最后再将映射的画面贴回包裹着虚拟摄像机的球体模型表面。这样便能够完成全景视频的制作。下面将介绍几种能实现全景视频制作的软件。

Unity3D

Unity3D 是由 Unity Technologies 开发的一个让玩家轻松创建诸如三维视频游戏、建筑可视化、实时三维动画等内容的多平台综合型游戏开发工具，是一个全面整合的专业游戏引擎。



图 1.25 Unity3D

MovieTexture 是 Unity 引擎内部自带，能够满足全景视频播放的组件，使用 MovieTexture 播放全景视频的效果较好，而且操作相对方便。但是，MovieTexture 只能播放 ogg 和 ogv，而且质量比较低。如果要转成高质量的视频，则文件尺寸比较大。

使用 MovieTexture 组件播放全景视频的方法是：

1. 在当前场景中新建一个 3D 的 Sphere 球体，将主相机的位置放在球体中心点处。
2. 将 ogv 或者 ogg 格式的视频文件放到 Resources 目录或者其子目录下，例如放在 Resources/Videos 目录下，通过以下代码来获取视频资源：

```
//不必带后缀名
```

```
MovieTexture tex = Resources.Load<MovieTexture>("Videos/video");
```

3. 获取 Sphere 对象上的 Renderer 组件，并将视频载入得到的 movTexture 纹理，传递给 Renderer 的 material.mainTexture 属性，并设置视频纹理的播放模式：

```
//设置当前对象的主纹理为电影纹理
```

```
_parenTrans.GetComponent<Renderer>().material.mainTexture = movTexture;
```

```
//设置电影纹理播放模式为循环
```

```
movTexture.loop = true;
```

4. 通过以上步骤便完成了视频的加入以及播放前的所有准备，那么接下来要做的就是开始播放、暂停和停止视频等操作，较为简单，分别调用 MovieTexture 的接口即可：

```
//开始播放
```

```
movTexture.Play();
```



```
//暂停播放
movTexture.Pause();
//停止播放
movTexture.Stop();
```

运行时,查看结果可以发现视频被赋值到材质球中作为纹理,通过旋转相机 Y 轴的角度,可以 360 度地观看视频,这就是全景视频播放的一个过程。

CINEMA 4D

德国 MAXON 公司出品的 Cinema 4D,是一套整合 3D 模型、动画与算图的高级三维绘图软件,一直以高速图形计算速度著名,并有令人惊奇的渲染器和粒子系统,其渲染器在不影响速度的前提下,使图像品质有了很大提高,可以面向打印、出版、设计及创造产品视觉效果。Cinema 4D 使用的是 Python 语言。

在 Cinema 4D 中可以使用全景摄像机来渲染全景视频:

```
Camera = c4d.BaseObject(5103)
Camera[c4d.CAMERAOBJECT_SPC_ENABLE] = 1
```

上述代码表示摄像机的 ID 为 5103。全景摄像机并没有专门构建出一个对象出来,而是在原有的标准摄像机的基础上扩展了全景功能。所以在调用了标准摄像机对象以后,第二行代码激活了全景摄像机选项。

全景摄像机有很多个参数,比如 FoV 辅助参数,映射参数等。我们主要对映射模式参数进行设置。全景摄像机的映射模式主要有两种:Lat-Long 映射和立方体映射,Lat-Long 映射也就是经纬度映射(EPR),它将球面上的各个点映射到了一个二维矩形面上,这是渲染全景图时经常采用的一个方法,这种方法能够比较方便地和传统视频处理及编码技术进行结合。但是这种方法会造成视野顶部和底部较为明显的失真,接缝明显。而立方体映射(六面体映射)则很好地解决了这个问题,将画面渲染到围绕观察点的六个垂直的面上,形成立方体包裹,保证了顶部和底部较小的失真。

VRay

VRay 是由 chaosgroup 和 asgvis 公司出品的一款高质量渲染软件。基于 V-Ray 内核开发的有 VRay for 3ds max、Maya、Sketchup、Rhino 等诸多版本,为不同领域的优秀 3D 建模软件提供了高质量的图片和动画渲染,方便使用者渲染各种图片。VRay 渲染器提供了一种特殊的材质——VrayMtl。在场景中使用该材质能够获得更加准确的物理照明(光能分布),更快的渲染,反射和折射参数调节更方便。VrayMtl 中可以应用不同的纹理贴图,控制其反射和折射,增加凹凸贴图和置换贴图,强制直接全局照明计算,选择用于材质的双向反射分布函数(BRDF)。

与其他的渲染器相比,VRay 的灵活性、易用性更好,有着焦散之王的美誉。VRay 还包括了其他增强性能的特性,包括真实的 3d Motion Blur(三维运动模糊)、Micro Triangle Displacement(级细三角面置换)、Caustic(焦散)、通过 VRay 材质调节完成 Sub-surface scattering(次表面散射)的 sss 效果、和 Network Distributed Rendering(网络分布式渲染)等等。VRay 特点是渲染速度快,目前很多制作公司使用它来制作建筑动画和效果图,就是看中了它速度快的优点。VRay 渲染器有 Basic Package 和 Advanced Package 两种包装形式。Basic Package 具有适当的功能和较低的价格,适合学生和业余艺术家使用。Advanced Package 包含有几种特殊功能,适用于专业人员使用。目前市场上有很多针对 3DSMAX 的第

三方渲染器插件，Vray 就是其中比较出色的一款。主要用于渲染一些特殊的效果，如次表面散射、光迹追踪、焦散、全局照明等，也可以使用 Vray 插件在 3DMAX 中渲染全景图片。Vray 是一种结合了光线跟踪和光能传递的渲染器，其真实的光线计算创建专业的照明效果，可用于建筑设计、教学等多个领域。

1.4 典型全景相机介绍

随着虚拟现实行业的兴起，原本只常见于机器人制造、工业、监控等领域的全景摄像机被越来越多地被应用到生活娱乐领域，大量新的厂商进入全景相机领域，市面上的全景拍摄器材日益丰富。目前，国外的 Nokia、Google、GoPro、Facebook、Jaunt、三星、松下、柯达、尼康等知名企业以及国内强氧、得图、Insta360、UCVR、兹曼等企业都在布局 VR 全景摄像机。

就目前的使用情况而言，全景摄像机按照镜头类别主要可以分为鱼镜头全景摄像机、多镜头全景摄像机两类。

(1) 鱼眼全景摄像机

鱼眼全景摄像机由单传感器配套特殊的超广角鱼镜头组成，工作原理在之前章节已经提到过，依赖图像校正技术还原图像。由于性价比高，此前这类产品在中占据主流份额，常被用于工业、监控等领域，比如考场、大厦的安保监控等。然而由于鱼镜头的特殊性，画面边缘畸变部分难以达到高清晰度，对于生活娱乐的市场而言，并不能满足需求。因此随着虚拟现实消费市场拓展，其所占市场比有所下降。



图 1.26 鱼眼摄像头

(2) 多镜头全景摄像机

多镜头全景摄像机是通过多个传感器配合特制的镜头组合实现全景功能的。由于多镜头拼接全景摄像机的各个传感器捕获的都是常规矩形图像，因此无需进行矫正操作。而其缺点在于大部分多镜头全景相机需要配套可实现画面无缝拼接的算法或软件，且在组装时对镜头

视场角与安装位置的设定都有严格要求。另外，由于多镜头全景摄像机相当于多个镜头+传感器的组合，因此成本自然会有所增加，性价比上低于鱼镜头摄像机，对于普通消费者而言略显昂贵。不过，多镜头全景摄像机的适用性相当广泛，可以支持各类生活、娱乐，甚至影视级别的全景视频内容拍摄，当前网络上的各类 VR 直播、VR 短片基本都是使用这类全景摄像机拍摄而成的。

以下将介绍几款典型的多镜头全景摄像机：

GoPro Omni

GoPro 是一款小型可携带固定式防水防震运动相机，被广泛应用在冲浪、滑雪、极限自行车及跳伞等极限运动上，被认为是“极限运动专用相机”。2016 年 4 月，GoPro 正式发布全景相机 GoPro Omni VR，这款相机采用了立方体结构，内部集成了 6 颗该公司的 Hero4 Black 摄像头。据 GoPro 方面称，该设备采用了特殊的技术，使得 6 颗摄像头之间能够实现“像素级别的同步”。



图 1.27 Gopro Omni

技术规格：

各摄像机分辨率/帧速率（球面分辨率）

- 2.7K 4:3 / 30 fps (7940x3970)
- 2.7K 4:3 / 25 fps (7940x3970)
- 1440p / 60 fps (5638x2819)
- 1440p / 50 fps (5638x2819)

视频格式：

H.264 codec、.mp4 文件格式，45Mb/s (Protune™ 60Mb/s)

音频格式：

48kHz，原始 PCM

媒体：

6 张 microSD 卡

连接/控制：

- 6 根迷你 USB 连接线，用于摄像机充电和卸载
- 1 个 Smart Remote，用于远程控制

软件:

Omni 特有阵列摄像机固件

电源:

- 6 颗可充电锂离子电池（规格：1160mAh，3.8V，4.4Wh）
- 1 个 12V 电源输入端（规格：2.5mm x 5.5mm，7A）

物理尺寸:

120mm x 120mm x 120mm

Surround 360

Surround360 是一个用于摄制和渲染 3D（立体）360 视频和照片的硬件和软件系统。其硬件部分的全景摄像机形状类似飞碟，共采用 17 部超高清摄像机环绕构成，并配有基于网络的软件，可在 360 度范围内捕捉图像和自动呈现。这款设备能够连续工作而不会出现过热现象，拍出的视频分辨率最高可达 8K。其全景视频可以在 Gear VR、Oculus 等虚拟现实头盔以及 Facebook 的应用上观看。这款相机的易用性在业界备受赞誉，其全部配件均可在市场上购买到。目前这款设备的售价为 3 万美元，为了降低售价，Facebook 方面已将 Surround 360 的硬件组装图和后期拼接软件放到了 Github 上，向全球开发者开放其成果，希望更多的开发者共同打造出更优质、更强大的全景摄像工具。



图 1.28 Surround 360

Jaunt VR ONE

Jaunt ONE 是专为摄制高质量立体 360° 电影虚拟现实体验而设计的专业级摄像系统。作为行业内首个电影级虚拟现实摄录一体机，Jaunt ONE 是 VR 创作先行者心目中的理想产品。经历了两年以上的高强度研究和开发，Jaunt ONE 不但拥有酷炫的外形，更拥有专为 VR 定制的光学系统，支持高质量全景音视频捕捉。

其主要特点为：

- 1) 24 目同步全局快门
- 2) 最高支持 120 帧/秒 (FPS)
- 3) 3D 全景覆盖
- 4) 单眼分辨率最高达 8K × 4K
- 5) 动态曝光调节
- 6) 单个模块自动/手动曝光调节
- 7) 远程相机操控及实时监测



图 1.29 Jaunt ONE

除了上述专用于拍摄全景视频的一体机，常见的攒机方案有：

多目 GoPro 方案

多目 Gopro 黑狗 4 方案是目前性价比最高，也是最受拍摄团队喜爱的方案。一方面，它可以根据需求任意组装，6 目、7 目、8 目、10 目、12 目甚至 14 目均可实现，没有任何限制。支架既可以从各平台上购买成品，也可以自己 3D 打印按需定制。另一方面，每台黑狗的成本在 3000 元左右，因此入门级设备 2~3 万即可拥有，且操作简单，成本低廉，上手门槛低。国内的天狗全景、莱瑞特、Upand 等团队都使用多目 Gopro 解决方案，它可以拍摄绝大部分非专业类全景视频和部分专业类全景视频。

强氧方案

强氧是目前国内较专业的全方位影像服务电商。其 2 代解决方案由 10 台 Drift Ghost-S 拼接而成，上两台，下两台，中间六台。Drift Foream Ghost-S 在长期运行、直播、散热、耗电量、稳定性等方面较 Go Pro 有一定的优势。其 3 代方案 Argus Panoptes Pro 在 2 代产品上更进一步，全景相机采用了环绕 8 颗，上方 1 颗摄像头的 9 目设计。感光元件采用的 M43 系统，支持拍摄 4K 60 帧的全景视频。强氧解决方案被广泛应用于各类大型活动直播，曝光率较高，官网商城的全景直播解决方案套餐价格为 12 万。

红龙拼接

如果需要拍摄影视级别的全景内容，则需要使用更加专业的解决方案。国内已有不少 VR 影视制作团队开始使用红龙拼接的方案。使用 4 台 RedDragon 面向四个方向进行拍摄，可以达到 24K 100FPS 的录制标准，也可以进行 30FPS 标准的 4K 直播。不过这套方案造价也

相当昂贵，价格近 200 万元，普通的消费和商业全景内容无需使用这类设备，否则处理巨大的素材源也将成为一大难题。其他基于红龙的 VR 电影解决方案也大致如此。6 目，7 目，8 目等等。

HeadcaseVR

2016 年,HeadcaseVR 团队正式公布了全新移动 VR 拍摄方案。这个创业团队来自好莱坞。专门从事 VR 电影拍摄工作。该方案主要采用 17 目 Codex Action Cameras。Codex Cam 有 12bit RAW 的记录体系和 13.5 档的高动态。采用 2/3 英寸的 CCD 传感器，单分辨率 1920×1080，23.98fps - 60fps 的帧率表现。头部尺寸为 45mm x 42mm×53mm，因而外观十分小巧，同时也配备专业的采集设备来实现录制。这个团队还定制了更适合移动 VR 视频拍摄的移动工具，虽然看上去像一台布满电池和采集器以及供电元件的轮椅，结构复杂又臃肿。不过这台移动设备解决了在 VR 视频拍摄中由移动产生的位移偏差及抖动问题。目前拍摄的负责人是 Marc Dando。

HypeVR

Hype 采用将 14 个 RedDragon 拼合的方式一步到位地实现了真正达到电影级 VR 设备的方案,该方案运用了激光雷达传感技术,由 Velodyne 公司设计,RedDragon 是 6K 级的。HypeVR 最终拼接完成可以达到 16K 90FPS 标准。

NextVR

跟 HypeVR 类似，NextVR 采用的也是 RED DRAGON 6K。但是他们选择了 6 台的拼接方案，三个方位，每个方位安放两台。虽然机器只有六台，但是价格依旧不菲。可以初步估算下：RedDragon 3 万美元一台，佳能 8-15mm f/4L 鱼镜头 1400 美元一个，监视器是 Red Pro，1600 美元一个，……。基于该方案，4K 直播可以轻松实现。

J2VR-极图全景

中国的团队 J2VR 目前也给出了经过几代更新后的电影级 VR 解决方案。采用 4 台 RedDragon 分别对四个方向进行拍摄采集。最终达到 24K 录制 100FPS 的标准。官方声称打造这个方案花费了 180 万重金，这让大众拭目以待。

Google JUMP

它实际上仅是一个支架。需要配合 16 台 GoPro 和专业软件才能实现全景视频采集和拼接。由于并没有设计向上或者向下的镜头，该方案只能在垂直范围内采集 120° 的图像。但是它的优势在于高性价比，花费约几万块就可以实现 360° 的 VR 视频拍摄。

本章参考资料：

- [1]王敏,周树道,张水平,黄峰.全景立体成像技术浅述[J].信息技术,2014(05):24-27+30.
- [2]肖潇,杨国光.全景成像技术的现状和进展[J].光学仪器,2007(04):84-89.
- [3]王琼华,王爱红.三维立体显示综述[J].计算机应用,2010,30(03):579-581+588.
- [4]韩伟.3D影视图像技术泛论[J].有线电视技术,2009,16(11):79-81.
- [5]丁剑飞,刘永进.三维立体显示技术综述[J].系统仿真学报,2008,20(S1):132-135.
- [6]郑华东,于瀛洁,程维明.三维立体显示技术研究新进展[J].光学技术,2008(03):426-430+434.

- [7] 冯传岗. 摘掉眼镜看 3D 浅谈影视图像的 3D 显示技术[J]. 数码影像时代, 2012(11):44-51.
- [8]https://www.chinatmic.com/360_Panoramic/show/92.html
- [9]<http://www.elecfans.com/instrument/581578.html>
- [10]<https://www.xzbu.com/8/view-4775061.htm>
- [11]<https://wenku.baidu.com/view/3424ef6fa9956bec0975f46527d3240c8547a15a.html>
- [12] Ishiguro H, Yamamoto M, Tsuji S. Omni-Directional Stereo[M]. IEEE Computer Society, 1992.
- [13] Shum H Y, Szeliski R. Stereo Reconstruction from Multiperspective Panoramas: IEEE, US6639596[P]. 2003.
- [14] 丁艳. 全方位视觉技术的综述[J]. 科技信息:科学教研, 2007(34):236-237.
- [15] 王健, 张振海, 李科杰, 等. 全景视觉系统发展与应用[J]. 计算机测量与控制, 2014, 22(6).
- [16]<http://baijiahao.baidu.com/s?id=1582008687252798978&wfr=spider&for=pc>
- [17]http://www.sohu.com/a/108274423_423652
- [18]<https://blog.csdn.net/linolzhang/article/details/54377060>
- [19]<https://wenku.baidu.com/view/9c9f3acded630b1c58eeb551.html>
- [20] 赵书睿. 全景图像拼接关键技术研究[D]. 电子科技大学, 2013.
- [21] 江铁, 朱桂斌, 孙奥. 全景图像拼接技术研究现状综述[J]. 重庆工商大学学报(自然科学版), 2012, 29(12):60-65+71.
- [22] 宋宝森. 全景图像拼接方法研究与实现[D]. 哈尔滨工程大学, 2012.
- [23]<https://baike.baidu.com/item/SIFT/1396275?fr=aladdin>
- [24]<https://baike.baidu.com/item/%E5%9B%BE%E5%83%8F%E8%9E%8D%E5%90%88/625475?fr=aladdin>
- [25]<http://www.goprochina.cn/cameras1/omni---%E4%BB%85%E5%8C%85%E5%90%AB%E5%A5%97%E7%9B%92/MHDHX-007.html>
- [26]https://www.sohu.com/a/109076657_411731
- [27]<https://github.com/facebook/Surround360>
- [28]<https://www.jauntvr.cn/technology/>