

第1章 绪 论

1.1 研究背景

随着光电成像技术的发展和便携式电子设备的普及,成像系统与模组被集成到各类移动设备中,已在手机成像、工业检测、机器视觉、生物医疗等领域取得了广泛应用。此外,包括可穿戴设备、机器人、物联网、虚拟/增强现实、人机交互等新兴的应用场景,也对与之匹配的成像模块的微型化、集成化提出了更高的要求。中国工程院信息与电子工程学部、中国信息与电子工程科技发展战略研究中心在北京发布了“中国电子信息工程科技发展十三大挑战(2022)”,总结提炼出中国电子信息工程科技当前的发展趋势和面临的重大挑战,其中包括信息领域的自主可控,实现超高速、高性能、低功耗、多功能、高密度光电子器件等。因此,新一代成像器件的研发将更注重降低单位带宽的传输成本,向智能化、小型化和集成化方向发展。

为了实现上述目标,人们在光学成像技术探索的道路上做出了不懈努力。一方面,随着亚微米、深亚微米和纳米技术工艺的不断发展和器件结构的改进,针对光强响应的图像传感器技术日趋成熟,图像传感器的分辨率、信噪比、动态范围等性能不断提高^[1-3]。另外,借助其他光学元件或附加额外的成像光路,可将其他维度信息(如相位^[4]、偏振^[5]、光谱^[6]等)转换为光强信息,从而被图像传感器所记录,实现对物理世界中光场信息的全方位感知。另一方面,随着移动互联网的兴起和计算机算力的提升,图像处理、计算机视觉、机器学习、大数据处理等研究领域取得了突飞猛进的发展。这些研究领域的相互碰撞和融合推动了计算成像的诞生和发展。计算成像从光传播和信息传递的角度对整个成像系统进行建模,并将计算能力引入成像系统中,旨在突破成像器件在信息记录上的瓶颈。由于计算成像将成像的重心由硬件转移到算法,不仅大幅降低了硬件成本,而且提高了成像系统的设计自由度,使成像系统在信息获取的维度和尺度上更加自由,为成像器件的智能化、小型化和集成化提供了有效的技术途径。

无透镜成像作为一种计算成像技术,其采用照明调制或者引入简单编码调制器件代替透镜对场景编码的方式实现图像信息采集,具有结构轻薄、易于构建的特点,受到广大科研人员的关注。传统基于透镜的成像模型通过设计透镜组,将物点一一映射到像传感器平面上完成信号的模数转换,透镜组的体积往往决定了成像系统的厚度。而无透镜成像打破了场景到图像一一对应的采样形式,摆脱了成像对透镜组的依赖,实现了成像系统的轻薄化。透镜成像与无透镜成像的比较如图 1.1 所示。

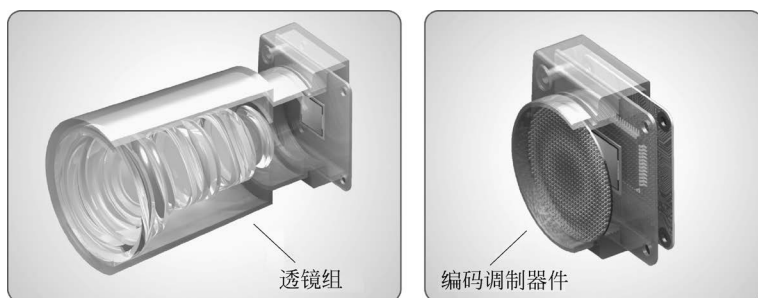


图 1.1 透镜成像与无透镜成像的比较

1.2 无透镜成像技术发展综述

根据成像系统对场景图像的编码调制方式的不同,无透镜成像技术可大致分为照明调制和掩模调制,如图 1.2 所示。照明调制又可分为空域调制和时域调制两大类;空域调制采用相干或部分相干光照明物体,通过衍

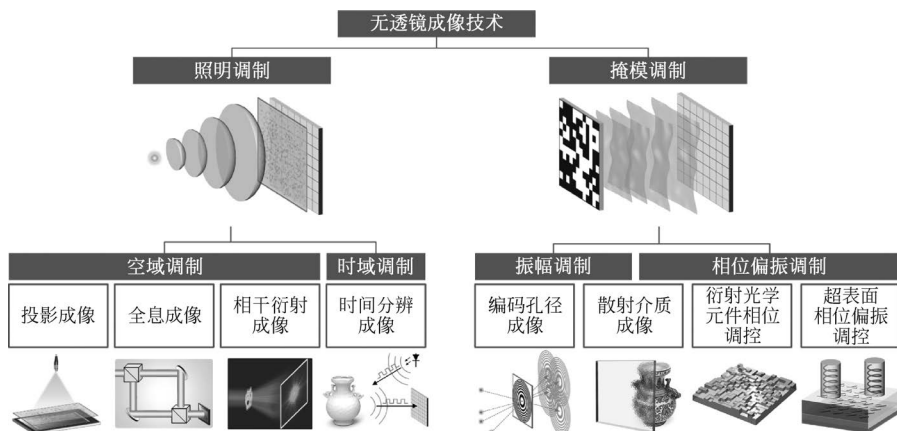


图 1.2 无透镜成像技术分类

射计算恢复物体图像,该方法需要主动照明,可实现对相位物体成像,但仅限微观尺度上对透明或半透明样品进行成像;而时域调制需要用到昂贵的成像设备(如条纹相机),仅适用于特定的成像领域。基于掩模调制的无透镜成像技术是本书研究的重点,掩模调制方式无须主动照明,可适用于显微成像和宏观物体成像,成像方式灵活,可实现多光谱成像及深度成像。

1.2.1 照明调制无透镜成像技术

基于照明调制的无透镜系统通过控制照明光与待成像物体的相互作用方式,利用照明光源的位置、相干性和脉冲时间等特性,采集一张或一组具有不同光照编码的图像,然后重建该编码图像即可实现无透镜成像。按照具体成像方式的不同,又可分为投影成像、全息成像、相干衍射成像和时间分辨成像。

(1) 投影成像。投影成像是基于照明调制的无透镜成像中最简单的一种成像方式,样品贴近像传感器表面放置,采用一束光强恒定且均匀的照明光源直接照明样品。当照明光源到样品的距离远大于样品到像传感器的距离,像传感器记录下样品的等比例投影图像。该方法无须任何重建算法,可以对半透明的微小粒子(如液滴和细胞)进行成像,被广泛运用于观察细胞或微生物的生长、运动以及其他特性的监测。2005年,Lange等设计了一种用于研究秀丽隐杆线虫的微流控投影成像装置^[7],测量了秀丽隐杆线虫的活性与环境温度间的函数关系。2008年,Ozcan等开发了基于投影成像的无透镜宽视场监测阵列平台^[8],称为“LUCAS”,如图1.3(a)所示;其视场面积与像传感器面积相当($37.25\text{ mm} \times 25.70\text{ mm}$),比同等分辨率光学显微镜视场大两个数量级,可以实现不同类型细胞的监测和计数。2011年,Zheng等通过改造像传感器^[9],基于投影式片上显微成像技术构建了一个智能培养皿平台,将整套系统置于培养箱内,实现了细胞生长以及胚胎干细胞分化过程的记录,如图1.3(b)所示。

由于投影式成像系统中不存在任何光学元件,其成像分辨率仅受到像传感器像素大小的限制。要想获得高分辨的投影成像,可采用小像素尺寸的像传感器进行成像,也可通过在像传感器上移动样品并采集多帧图像进行融合,实现亚像素分辨率成像。2010年,Zheng等采用像素超分辨算法,对裸藻、微球以及侵袭内阿米巴囊肿进行成像^[10],取得 $0.75\text{ }\mu\text{m}$ 的极限分辨率,远远超过像传感器像素尺寸 $3.2\text{ }\mu\text{m}$ 的限制。另外,和其他光学超分辨技术的结合也能够进一步提高投影成像的分辨率。2018年,Tønnesen等提出基

于超分辨率投影成像(SUSHI, Super-resolution SHadow Imaging)的新型脑组织细胞显微术^[11],如图 1.3(c)所示,将投影成像与超分辨荧光成像技术相结合,实现脑组织特定区域的所有脑细胞同时成像,分辨率最高可达十纳米量级。

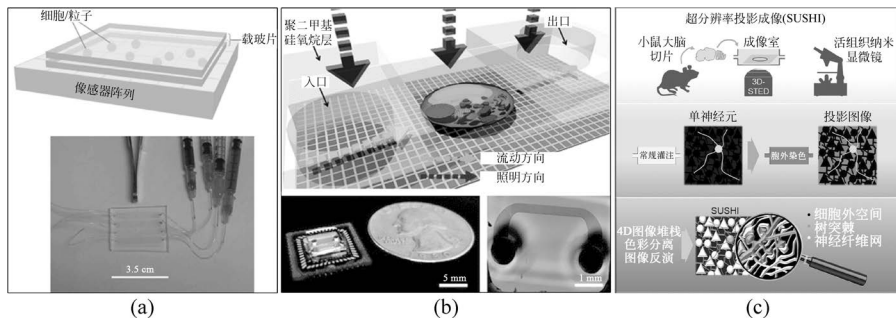


图 1.3 投影式无透镜成像系统

- (a) 无透镜宽视场监测阵列平台 LUCAS; (b) 亚像素光流显微镜;
(c) 超分辨率投影成像(SUSHI)流程图

在投影式成像结构中,需要保证样品和像传感器的距离足够小,通常在 $500\text{ }\mu\text{m}$ 以内^[12]。若增大样品与像传感器的距离,受到样品调制的照明光波将发生衍射,并随着传播距离的增大逐渐偏离样品的投影图像。此时照明光源若为非相干光源,样品图像则会变得模糊不清;而采用相干光源进行照明,样品的衍射图案与其自身光强透过率存在数理关系,可以通过衍射成像技术恢复样品信息,此类方法也被称为“无透镜全息成像”或“相干衍射成像”,相关内容将在下面详细讨论。

(2) 无透镜全息成像/相干衍射成像。无透镜全息成像系统使用相干或部分相干光源照明待测样品,在像传感器上获得样品的衍射图案。与投影成像类似,无透镜全息成像通常用于芯片实验室的显微成像。全息成像能够编码光波的复振幅信息,因此在图像重建时能获取比投影成像更为丰富的信息。光源的相干性是全息成像中的一个关键参数,虽然激光能够提供良好的空间和时间相干性,但是容易受到相干散斑噪声的干扰。发光二极管(LED, light emitting diode)可以根据光源的大小、带宽和到样品的距离来调节空间相干性^[13-14],因此在实际应用中被广泛采用。

由于像传感器本质上记录的还是光强信息,需要通过重建算法进一步恢复样品的复振幅。一种重建的方法是将像传感器上的测量图案视为样品的同轴全息图^[4],即测量图案是由照射到样品上散射的物光与未受样品干扰的

参考光之间干涉产生的强度图案,然后通过数字反向传播方法(如角谱法)重建样品的复振幅^[15]。然而同轴全息图中不仅包含原始物光波的复振幅,还包含样品的散射光强以及物光波的共轭光波,采用数字反向传播方法会在重建图像上叠加背景噪声和孪生像,造成成像质量的下降。在一些成像应用中^[16-17],所观测的目标样品尺寸较小且相对稀疏,背景噪声和孪生像不会对重建图像的质量造成较大影响,简单的反向传播方法就可以满足这些应用的需求。对于非稀疏样品而言,重建图像质量则会受到背景噪声和孪生像的干扰。特别是当样品具有复杂的振幅和相位变化时,其散射光强将远远大于与参考光之间的干涉光强,重建图像将淹没在背景噪声中而无法分辨,因此需要采用其他方法实现高质量图像重建。

一种常用的方法是把测量图案看作相干衍射图案,采用相位恢复算法实现样品相位和振幅信息的提取。传统相位恢复算法需要预先获取关于样品的先验信息(如样品的振幅^[18]、支撑区域^[19]等),通过在物平面和测量平面之间迭代施加先验约束^[20-21],实现目标样品的重建。然而在实际应用中,样品的振幅和支撑信息难以获取,此类算法仅能对纯相位或稀疏且具有清晰边界的样品进行成像。而采用多次测量的方式能提供附加的约束信息,可以用来重建(如组织切片等)高密度样品图像。例如,Greenbaum等采集了多张不同衍射距离的强度图片^[22],通过求解光强传输方程(TIE, transport of intensity equation)实现了对人类乳腺癌的组织切片、巴氏涂片以及全血涂片的成像,实验装置及图像重建流程如图1.4(a)所示;Luo等采集了多张不同角度照明下的测量数据来重建人体乳腺癌组织切片^[23],如图1.4(b)所示。

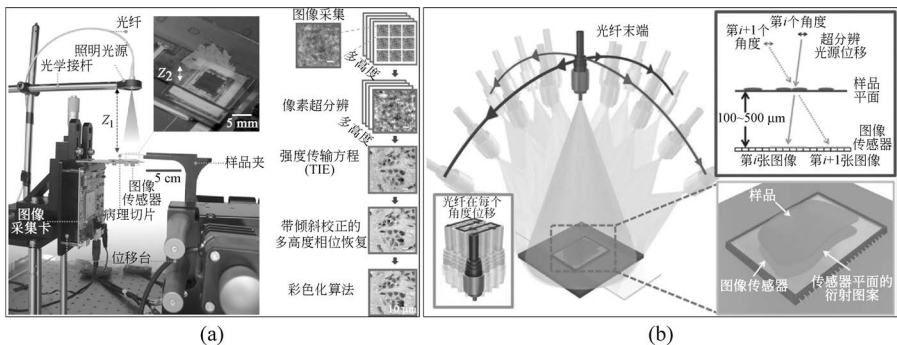


图 1.4 无透镜显微成像系统

(a) 基于多平面的无透镜显微成像装置及图像重建流程; (b) 基于多角度照明的合成孔径无透镜显微成像装置

由于全息成像将三维物空间的复振幅分布编码成二维强度图,因此可将压缩感知相关理论引入图像重建过程中,使得全息重建质量得到显著提高。压缩感知是由 Candès、Tao 和 Donoho 在 2006 年前后提出的一种信号处理框架^[24-25],其目的是通过构造特定的编码方式,实现用较少的信号采样来重建原始信号。2007 年,Latychevskaia 等指出衍射传播过程自身就是一种有效的编码方式^[26],利用传播过程的物理约束即可实现孪生像的消除。2009 年,Brady 等首次提出压缩全息的概念^[27],在理论上证明了全息衍射的编码方式满足压缩感知的“不相关性”条件,并实现了仅用单幅二维全息图重建蒲公英的三维层析图像。2011 年,Hahn 等利用压缩全息技术实现了对两只活体剑水蚤的层析图像采集^[28]。然而,压缩全息重建同样面临着对样品稀疏性的限制,无法直接用于重建稠密的生物样本。为了弥补这一不足,可将压缩重建方法中的稀疏表征与多次测量方法相结合,从而减少测量次数,提高成像速度^[29]。压缩重建也可用于无透镜多光谱成像,对多波长照明获得的图像进行解复用^[30]。

(3) 时间分辨无透镜成像。时间分辨无透镜成像装置由主动照明的脉冲光源和具有时间分辨能力的像传感器组成,通过记录光在光路中的反射时间差进行成像。Kirmani 等通过仿真验证了采用脉冲光源和小型阵列时间分辨像传感器可实现无透镜成像^[31]。Satat 等提出了一种时间分辨无透镜成像框架^[32],该框架采用特定的照明图案以及像传感器布局,利用压缩感知原理进行重构;该方案与传统的单像素相机相比,时间分辨探测器需要的测量次数减少为原来的 1/50。目前时间分辨无透镜成像对硬件要求较高,需要用到脉冲激光源进行照明,并采用单光子雪崩二极管 (SPAD, single photon avalanche diode) 阵列或昂贵的条纹相机^[33]进行探测。

1.2.2 掩模调制无透镜成像技术

虽然基于照明调制的无透镜成像系统结构简单紧凑,但由于没有额外的光学调制器件调节成像视场和放大率,其视场和分辨率与像传感器面积及像素尺寸直接相关。另外由于照明光源具有高度相干性,照射到粗糙的物体表面会引起随机散斑,破坏光波对物体的编码。因此,基于照明调制的无透镜成像主要应用场合局限于透明或半透明的生物样品的显微成像领域。基于掩模调制的无透镜成像系统则是通过在像传感器前放置一片光学掩模,实现对场景图像的编码调制。相较于照明调制方式,掩模调制成像具备更多的成像自由度,通过设计特定的编码方式和图像重建算法,可以灵活地应

用于各种场景,如三维成像^[34-35]、多光谱成像^[36-37]、重聚焦成像^[38-39]、显微成像^[39-40]等,这展示了掩模调制的无透镜成像方式在高通量、多功能成像方面具有应用潜力。

小孔成像是掩模调制成像最简单的一种形式。早在公元前 400 多年的战国时代,哲学家墨翟便在《墨经》里记载了有关小孔成像现象的描述。同时期,古希腊哲学家亚里士多德也有关于小孔成像的论述,但并未揭示小孔成像的原理。直到约 1400 年后,阿拉伯科学家海什木意识到人之所以能看见物体,是因为光线照射到物体表面,再从物体表面反射进入眼睛所导致的,而非之前人们认为的“人能看见物体是靠眼睛发射出的光线被物体反射的结果”,并通过复现小孔成像实验来证明自己的观点。他在黑暗房间里的窗户上开一个小孔,这样窗外的世界就会在窗户对面的墙上投射出一个颠倒的图像,如图 1.5(a)所示^[41]。海什木将其关于光学的理论和思想记录在了著作《光学之书》(*Book of Optics*)中,随着该著作传入欧洲,其对欧洲的科技革命产生了深远影响。自 15 世纪开始,欧洲出现基于小孔成像原理的暗箱,可以把影像投影在屏幕上。然而暗箱所形成的影像只可观察,不可记录,仅被艺术家用作绘画的辅助工具。1550 年,意大利的卡尔达诺将双凸透镜置于原来的小孔位置上,成像的效果比暗箱更为明亮清晰。1839 年,法国的达盖尔利用其发明的“银版摄影法”制成了第一台实用的银版照相机(见图 1.5(b)^[42]),实现了影像的长期保存,就此揭开了现代摄影术发展的序幕。

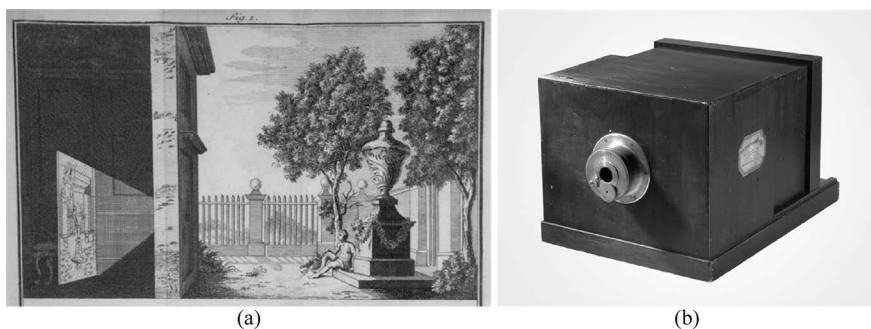


图 1.5 小孔成像的发展

(a) 暗室中的小孔成像现象; (b) 达盖尔所设计的暗箱相机

小孔成像因其光通量较少,需要长时间曝光才能获得令人满意的图像,目前仅在原理展示实验和长曝光摄影中有所应用。而编码孔径成像技术基

于小孔成像的原理做了进一步拓展,用一个包含多个透光孔的掩模版代替了单个小孔进行成像,解决了小孔成像曝光时间长、信噪比低的问题。编码孔径成像技术最初是为 X 射线和伽马射线成像而发明的,这类短波长射线具有极高的光子能量,能直接穿过透镜而几乎不发生偏折,因此无法采用透镜对其进行聚焦。编码孔径技术仅利用光的直线传播定律,通过在空间上对入射光的透过和吸收两种状态进行二元调制,巧妙地记录下场景图像的信息。

与透镜成像相比,编码孔径成像采集到的图像不再是“所见即所得”,而是一幅包含场景光强信息的编码图像。假设在黑暗背景中存在一个点光源,像传感器将采集到编码掩模的投影图像。如果改变光源的入射角度,那么像传感器上的投影就会移动;如果改变光源的深度,那么投影的大小就会改变。场景图像和像传感器采集图像之间的关系可用一个线性系统来表示,系统参数由掩模版图案和位置决定,使用适当的算法反演这个系统即可恢复场景的图像。因此,掩模版的设计在编码孔径成像中起着重要的作用。理想的掩模版能最大化进光量,同时提供良态的系统传递函数,使得图像重建更加稳健。常用的编码掩模包括波带片编码掩模^[43]、随机孔径阵列^[44]、均匀冗余阵列(URA, uniformly redundant arrays)^[45]以及改进的均匀冗余阵列(MURA, modified uniformly redundant arrays)^[46]等。

最早被提出用于编码孔径的掩模版是非涅耳波带片(FZP, Fresnel zone plate)。1950 年, Rogers^[47]注意到 FZP 和点源全息图之间的相似性,认为全息图是具有复杂图案的广义波带片。1961 年, Mertz 和 Young^[48]提出将 FZP 用作编码孔径,在非相干照明下可以得到类似伽博全息图的编码图像,然后通过光学重建的方式解码图像;该方法也被称为“波带片编码成像”(ZPCI, zone plate coded imaging),被广泛应用于天文学^[49]、核医学^[50]和激光惯性约束聚变^[51]等领域。

波带片编码成像包含记录和再现两个过程,如图 1.6(a)所示:第一步,用胶片记录下从辐射源发出的射线通过波带片产生的投影图案。辐射源中的每个点都在胶片上留下一个波带片的投影,投影的大小和位置代表了其对应源点的空间位置。第二步,用低功率、可见光波段的激光束照射经过处理后的胶片,每个波带片的投影将入射的激光聚焦到衍射受限的光斑上,实现源图像的再现。ZPCI 技术理论上需要采用透过率连续变化的伽博波带片(GZP, Gabor zone plate)作为编码孔径,由于 GZP 难以制作,所以从 ZPCI 技术提出以来人们都采用 FZP 作为编码孔径,如图 1.6(b)所示。FZP 具有多个焦点,而 GZP 仅包含一对共轭焦点,因此采用光学方法重建图像时

FZP 的高阶衍射级次会在主焦平面上形成离焦图像,造成重建图像质量下降。1992 年,Beynon 等提出一种二值化的 GZP^[52],大大降低了 GZP 的制作难度,同时保留了 GZP 的聚焦特性,得到了较为广泛的应用。

Beynon 认为二值波带片环带的宽度不仅可以沿径向发生变化,也可以随着方位角的不同而变化。通过调整不同方位角的二值化透过率,使其沿圆周上的积分与 GZP 的对应半径上的透过率一致,即可实现 GZP 的二值化。具体地,可将一个波带片划分为若干个扇区,调整每个扇区上不同半径对应的圆弧透光部分占整个弧长的比例,使其与 GZP 相同,如图 1.6(c)展示了一个具有 16 个扇区的二值化 GZP。这样生成的波带片也被称为“汉普顿宫”波带片,因为其图案与英国汉普顿宫的天文钟外形很相似。这种波带片具有明显的辐条状结构,会导致重建图像中产生伪影。为了解决这一问题,Kirk 设计了一种旋转式二值化的 GZP,将相邻环带交错旋转半个扇区,消除了辐条状结构,如图 1.6(d)所示。可以预见,随着扇区数量的增加,二值化 GZP 将越来越接近真实的 GZP。

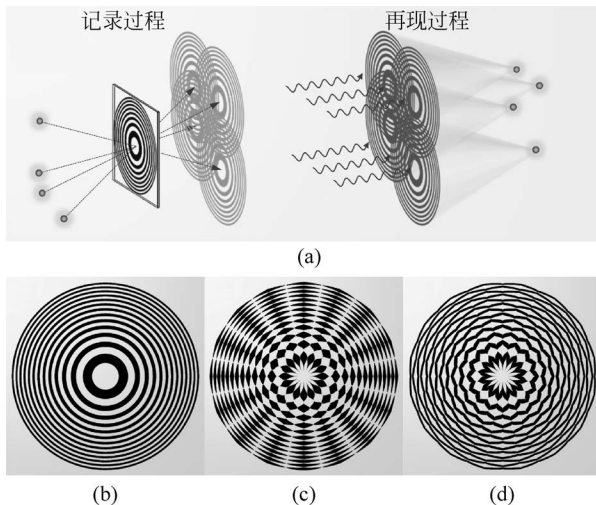


图 1.6 波带片编码成像原理图

(a) 波带片编码成像的记录过程和再现过程; (b) FZP 图案; (c) 二值化 GZP 图案; (d) 旋转式二值化 GZP 图案

另一类广泛应用的掩模版是由多个小孔组成的孔径阵列。Dicke^[53]和 Ables^[54]于 1968 年分别独立提出了用于 X 射线和伽马射线成像的随机孔径阵列。随机孔径阵列本质上是对小孔成像原理的拓展。小孔成像的分辨

率随着孔径尺寸的增大而降低,为了获得高分辨率,小孔的尺寸应当尽可能地小。然而缩小孔的尺寸会导致通光量减少,成像灵敏度也会随之降低。显然,在小孔成像模式下,分辨率和灵敏度是一对相互制约的参量。随机孔径阵列的解码,通过增加小孔的数量提高通光量,同时保持分辨率依然与每个小孔的尺寸相当。基于孔径阵列的编码图像的解码方法是用掩模图案与编码图案进行卷积,卷积图案的峰值表征了点源的强度和位置,从而再现辐射源图像。影响重建图像质量的一个重要特性是掩模图案的自相关函数,理想的自相关函数应当接近冲激函数。随机孔径阵列的自相关函数并非最优,其旁瓣噪声过高且不均匀,对扩展源的成像会导致这些旁瓣在背景中产生伪峰值信号,干扰成像结果。

1971年, Golay^[55]提出了一类具有近乎完美成像特性的编码掩模图案,称为“非冗余阵列”(NRA, non-redundant array)。NRA图案的自相关函数仅可能出现三个离散值:中心峰值、围绕中心峰值一定半径内的平台值,以及其余的零值或者非零均值。但 NRA图案中透光孔的数量比较少。1978年, Fenimore等提出均匀冗余阵列(URA, uniform redundant array)^[45],其自相关函数只有两个值:中心峰值与其他非零均值。相对于随机孔径阵列和 NRA而言, URA的最大优点是它具有均匀的旁瓣,因此解码图像的背景噪声可以通过减去一个常量噪声值的方法消除。但 URA仅能生成矩形的编码孔径,在圆形孔径的成像设备上使用略有不便。1989年, Gottesman^[46]在 URA的基础上提出了改进的均匀冗余阵列(MURA, modified uniformly redundant arrays),它兼顾了 URA阵列的所有优点,可以排列成六边形(HURA, hexagonal URA)^[56],使得加工与使用都更为方便。目前,基于 MURA和 HURA的编码孔径已成功应用于工业伽马射线相机和高能天文望远镜,如美国 RMD公司研制的小体积伽马射线相机 RadCam2000TM,以及国际伽玛射线天体物理实验室(INTEGRAL, International Gamma-Ray Astrophysics Laboratory)卫星平台所搭载的光谱仪,如图 1.7 所示^[57-58]。

尽管当前编码孔径成像技术在 X 射线、伽马射线等短波长成像领域已有成熟的商业化应用,但相关成像方法并不能直接应用到可见光波段。其原因在于二者观测目标不同,对成像质量的要求也不同。在短波长成像领域,辐射源往往是空间稀疏的,成像的目的是获得辐射源的位置和大小,因此成像分辨率是相对较低的。要想提高成像分辨率,则需要缩小掩模版的透光孔尺寸(特征尺寸),然而在可见光波段缩小特征尺寸会引起显著的衍