

多相机系统:成像增强及应用

郭珮瑶, 蒲志远, 马展*

南京大学电子科学与工程学院, 江苏 南京 210023

摘要 单相机是生物单目视觉的直观模拟,而自然界中的生物主要以双目及多目视觉系统实现场景的精确感知。受此启发,研究人员认为可以应用多相机阵列来增强成像质量。通过不同相机的不同传感器和拍摄条件,实现多维多尺度图像/视频的采集,而后利用相机间的匹配实现不同特征域上的计算融合,从而达到重建场景在尺度或维度上的增强。本文以宽视场超高清成像、时空高速视频采集、高动态范围成像及低照度成像为例,详细介绍了已有多相机系统在成像增强方面的应用,归纳总结多相机系统在提高成像时空分辨率、扩大成像视场、丰富成像动态范围等方面的优势。

关键词 成像系统;多相机系统;计算成像;成像增强;时空分辨率;动态范围

中图分类号 TP751

文献标志码 A

doi: 10.3788/LOP202158.1811013

Multi-Camera System: Imaging Enhancement and Application

Guo Peiyao, Pu Zhiyuan, Ma Zhan*

School of Electronic Science and Engineering, Nanjing University, Nanjing, Jiangsu 210023, China

Abstract A camera mimics the biological vision system for acquiring natural scenes. Human beings and other animals are mostly equipped with binocular stereo or multi-eye vision systems for high-fidelity scene acquisition, which motivates us to develop a multi-camera system to enhance the imaging capacity. Multi-dimensional and multi-scale image/video acquisitions can be realized via heterogeneous sensors and shooting conditions of different cameras, and then leveraging matches across cameras to realize computational multiscale feature fusion for final enhanced reconstruction in respect of scale or dimensionality. In this article, we exemplify the multi-camera system in the applications of wide-field ultrahigh-definition imaging, high spatiotemporal video acquisition, high dynamic-range and low-light imaging enhancement to demonstrate its advantages in improving the imaging capacity, e. g., increasing imaging spatiotemporal resolution, expanding the field of view and extending imaging dynamic range.

Key words imaging system; multicamera system; computational imaging; imaging enhancement; spatiotemporal resolution; dynamic range

OCIS codes 110.1758; 040.1490

1 引言

自相机发明以来,提升其成像能力一直是业界追求的目标,如高动态范围(HDR)采集,超高分辨率或十亿像素(gigapixel imaging)成像,时空高速(high spatiotemporal resolution imaging)或高速

(high speed)录像,低照度(low-light imaging)成像等。但由于光的衍射极限、传感器物理参数(如采集速率和传感器动态范围)、环境噪声等的限制,单相机捕获的信息受限,制约了它的成像性能,很难同时兼顾在不同维度及不同尺度等多方面的高质量重建。例如,一般的高速视频采集过程需要牺牲空间

收稿日期: 2021-05-31; 修回日期: 2021-07-05; 录用日期: 2021-07-20

基金项目: 国家自然科学基金(62022038, U20A20184)

通信作者: *mazhan@nju.edu.cn

分辨率来保证即时的信息存储,而宽视场成像,如全景(panorama)拍摄,则会以时间维度换取空间视角信息。

阵列相机(array camera)通过多个相机的组合来获取更丰富的不同维度和尺度的场景信息,可以帮助实现更全面的图像重建^[1],突破单相机成像的局限。应用多相机系统(multi-camera system)实现高质量成像的核心是如何利用不同的相机对场景进行精准采样,以及如何有效融合已捕获的信息来帮助重建。近 20 年来,蓬勃发展的计算成像技术(computational imaging)已成功实现了对光学成像系统的建模,并利用数学手段从观测图像中逆向求解出了场景目标的高质量、高维度物理信息。目前已有的工作提供了不少利用多维多尺度特征进行计算采样与重构的思路,包括基于先验规则(rules-based)的方法^[2],以及当前的研究热点——基于深度学习(deep learning-based)的策略^[3]。同时,在产业中也涌现出了基于多相机成像模组的智能摄像设备,如:Light 公司在 2015 年宣传的 L16 产品(<https://www.amazon.co.uk/Light-B077RP4MHN-light-lens-camera/dp/B077RP4MHN>,该产品通过 16 个焦段不同的相机实现了变焦段超高清拍摄),Insta360 公司推出的消费者级别的 Pro 2 产品(https://store.insta360.com/product/Pro_2,该产品利用多相机环形架构实现实时全视角 4K 视频推流),Raytrix 公司上线的 R42 系列(<https://raytrix.de/wp-content/uploads/software/R42.pdf>,R42 系列通过多透镜阵列实现了千万像素级别的 3D 光场视频采集),安科迪公司的 Mantis 系列(http://www.aqueti.cn/mantis70_0601.pdf,该系列结合 18 个窄视场角相机与 1 个广视场角相机实现了高达 1.07 亿像素的广域视场高清视频拍摄)。此外,随着微型处理器计算能力的提升,多相机成像系统已成为移动影像设备(如华为、小米、苹果等系列智能手机)的标配。应用多相机系统探索多维多尺度的计算融合来实现高质量成像也成为这些智能设备巨头争夺用户的必要手段。

本文立足多相机成像增强,对已有的相关研究进行概述总结,深入讨论其在宽视场超高清成像、时空高速视频采集、高动态范围成像、低照度成像中的应用,细述各方案中利用多维多尺度特征进行计算采样与融合重构的策略。同时,本文主要关注多相机系统在成像视场范围、成像时空分辨率以及动态范围等方面的增强表现,对于其

在谱段融合^[4]、相位增强^[5]、偏正感知^[6]等方面的应用留待后续归纳。

本文首先介绍多相机成像增强的基本原理,引出多相机成像融合增强的问题,继而针对常见的多种应用案例分章节进行深入讨论,最后对多相机系统在成像增强方面的作用进行总结,并展望了未来小型化、轻便化的多相机系统结构在更多场景下的普及应用。

2 多相机成像增强的基本原理

光信号作为物理世界信息感知的重要载体,其特征涉及空间、角度、时间、动态范围与极化方向等多个维度,从而承载着不同的场景特性^[7]。Adelson 等^[8]于 1991 年提出用七维全光函数 $L(x, y, z, \theta, \phi, \lambda, t)$ 来描述空间中任意位置 (x, y, z) 、任意角度 (θ, ϕ) 、任意波长 λ 、任意时刻 t 的光线。而经典成像模型下的单相机对曝光时间内来自不同角度的光线进行沿离散频谱通道(一般为 R、G、B 通道)的响应积分后投影至二维焦平面,最终获得七维全光函数的二维投影子空间信号,仅在空间维度上保留了高分辨率信息。当下市场上传感器的规模尺寸、像元大小、响应灵敏度等存在一定限制,进一步制约了单相机光学成像在空间维度、时间维度、特征维度与探测灵敏度等方面的表现。随着成像电子学以及计算机数据处理技术的发展,众多研究者尝试将信号处理计算手段引入到成像过程中(部分基于深度学习的典型计算光学成像技术如图 1 所示^[9]),对光信号进行低维优化耦合采集,而后通过计算重构获取真实光信号的高维高分辨率信息^[10-11],以此进行光学成像增强。例如,采用微透镜矩阵采集光场来提升角度分辨率^[12],利用掩模调制来提升光谱的分辨率^[13]等。

人类及绝大多数哺乳动物自然演化出了双目视觉系统,而螳螂、蜻蜓等则是多目复眼系统^[14],以满足不同的生存需求。多目生物视觉系统可以在长距离、宽视场、极大动态范围以及暗光等条件下实现对场景的良好感知。受上述自然界生物视觉多目采集的启发,研究人员积极探索多相机组合(阵列)来实现真实场景的高质量成像重建。

相较于单相机系统,多相机系统成像可以丰富捕获信息在空域、时域、频谱等维度上的特征,且各相机间对场景各维、各尺度的采样并不一定是均匀分布的(具体见图 2^[15])。在进行多源信息成像重建过程中无疑要引入计算融合来实现相机间多尺度

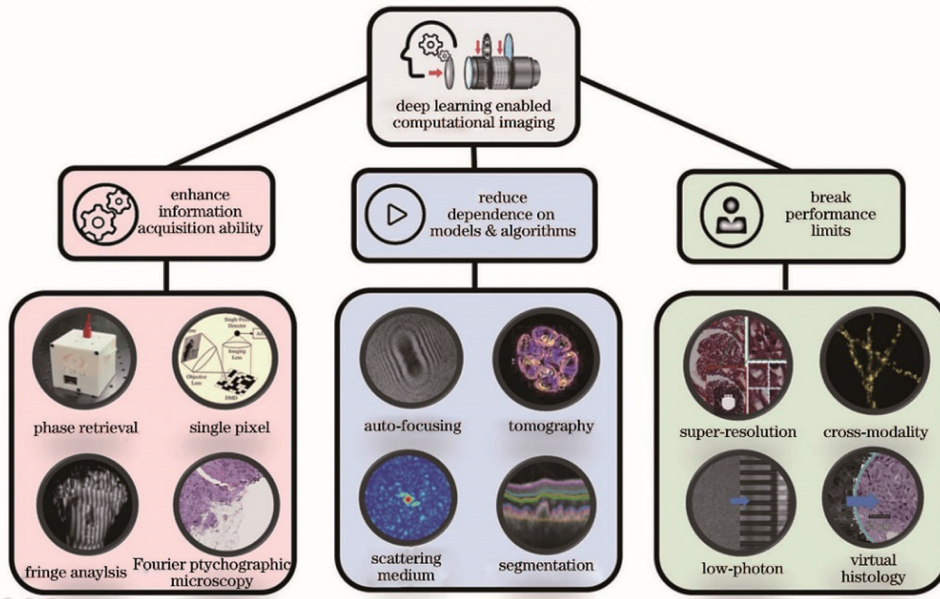


图 1 基于深度学习的典型计算光学成像技术^[9]

Fig. 1 Typical deep learning based computational imaging techniques^[9]

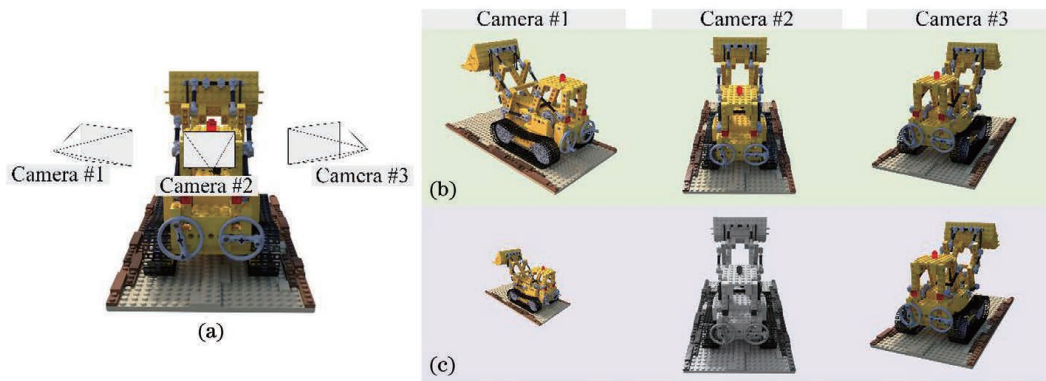


图 2 多相机系统对模拟场景的多维多尺度采样说明^[15]。(a) 模拟场景下,三相机视点示意图;(b) 三相机对场景各维各尺度均匀采样,即三相机的成像控制参数一致;(c) 三相机对场景各维各尺度非均匀采样,即三相机的成像控制参数不一致,例如 #1 相机采用低空间分辨率,#2 相机未保留色彩信息

Fig. 2 Illustration of various dimensional and multi-scale data acquisition with multi-camera systems in the toy model^[15]. (a) The illustration of three cameras' viewpoints in simulation scene; (b) uniform sampling in the multi-dimensional scene features, namely same imaging control setting for all cameras; (c) non-uniform sampling in the multi-dimensional scene features, namely various imaging control settings among cameras. For example, camera #1 captures at the low spatial resolution and camera #2 doesn't record color information

特征的交互补偿。具体来说,阵列中各相机的最终成像质量(分辨率、帧率、亮度等)受控于各相机内部光敏元件的光电转换效率以及相机的快门、光圈、感光度(ISO)等参数的设定,在不同的相机上应用不同的成像控制参数可以实现多相机阵列对同一场景的多维多尺度非均匀采集。而后通过不同相机间的配准实现相机间信息的计算融合,就可以实现高质量图像/视频的重建。

利用多相机系统进行成像增强的工作流程一

般为:

- 1) 利用多个相机对场景进行不同分辨率、不同特征维度的采样,优化光信号的低维耦合采集;
- 2) 基于空域/特征域对捕获的多源信息进行配准,以获取不同相机间每个像素处各维度特征的对应关系;
- 3) 利用该对应关系实现不同相机间的特征融合补偿,以实现各维度空间内的采样增强。

针对不同的多相机系统成像增强任务,多相

机的耦合采集方式各异:为获取高速超清视频,可使用高空间分辨率与低时间分辨率的相机组合,或者使用低空间分辨率与高时间分辨率的相机组合;通过设置各相机间不同的曝光时长,捕获场景在更宽动态范围下的信息。不同相机的采集条件相异,因此如何实现多相机之间的精确配准对于多源信息的交互融合影响重大。经典的匹配方法有单应性矩阵变换^[16]、光流矫正^[17]等。此外,新近出现的基于深度学习的匹配方法,如局部单应性匹配^[18]、可变形卷积^[19]等,也逐渐得到广泛应用。对于多源信息的交互融合,一般分为像素级别、特征级别以及决策级别三类,其中:像素级别的融合直接使用经过不同源图像配准后的像素进行处理得到最终的融合图像;特征级别的融合先对配准后的多源图像进行特征提取、融合,再估测出最终的图像;决策级别的融合一般利用目标检测、识别等高阶图像任务对多源图像的决策输出进行融合,从而得出最终的决策结果。近年来,卷积网络凭借其对于高维特征更好的表征能力^[20],逐渐代替传统的拉普拉斯变换、小波变换^[21]或引导滤波^[22]等方法来实现多源信息间的高质量融合。

下文将通过不同的实际案例逐步介绍多相机在宽视场超高清成像、时空高速视频采集、高动态范围成像和低照度增强等方面应用的主要架构及性能。

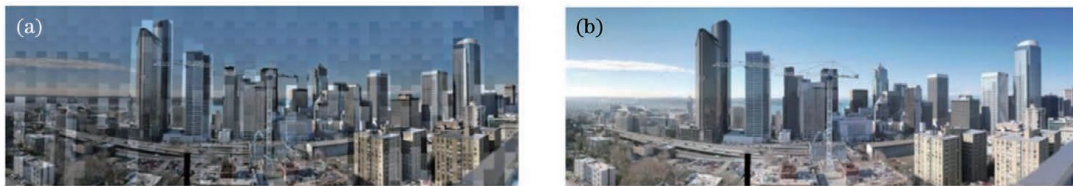


图 3 单相机扫描拍摄及拼接结果^[24]。(a)直接拼接的结果;(b)对(a)进行亮度对齐和色调映射处理后的结果

Fig. 3 Mosaic results acquired with single camera through scanning^[24]. (a) Captured images with only geometric alignment; (b) result after radiometric alignment and tone mapping

传统阵列相机依赖于精密的光学、机械设计,大部分需要繁琐的校准以保证拍摄精度,同时需要处理体量庞大的数据流,给系统设计带来了较大挑战。可实现宽视场超高清成像的阵列相机系统最初设计为规则排列的相机阵列,其设计遵从均匀感知原理(uniform sensation principle),利用计算成像技术同时保证高分辨率和大视场。杜克大学在此基础上提出了多尺度设计的概念^[25],使得此类阵列相机系统得以小型化。清华大学最新的系统解除了子相机均匀排列的限制,在多尺度阵列相机的基础上进一步加入了模块化设计,所设计的非结构化相机系统可以根据实际场景分布调整局部相机分布^[26]。

3 宽视场超高清成像

宽视场超高清成像旨在生成同时具有大视场和高分辨率的图像或视频,在智慧城市、卫星侦察、医学显微等领域均有应用。利用宽视场超高清成像设备能够对目标场景进行广视域成像及分析处理,进一步推动基于此类数据的社会行为分析。

在传统的相机设计中,镜头速率(lens speed)和视场随着镜头孔径的减小而增大。即:为了获取更大的视场,需要减小系统焦距(孔径减小,传感器尺寸不变),从而使得像素密度降低,图像分辨率下降。因此,传统的单孔径成像无法同时实现宽视场和高分辨率成像^[23]。一种简单的获取宽视场超高清图像的方式是利用高分辨率相机按照一定路径扫描拍摄后再拼接,但是该方式获取的首帧到尾帧间有较大时间差,每个拍摄时刻都有大面积成像盲区,无法正确处理场景中的运动物体^[24]。Kopf等^[24]用一台高分辨率佳能相机旋转捕捉不同位置曝光最佳的图像,然后对图像进行匹配拼接得到了十亿像素级宽视场高清图像(见图3^[24])。此方案中上百张图像的拼接合成所需处理时间过长且应用限制较大。而阵列相机可近似看作是一个大合成孔径相机,有效提高了相机系统的孔径面积,克服了单孔径相机的光学局限性。

本章归纳了基于传统阵列相机及多尺度阵列相机的宽视场超高清典型成像方法,分析了各系统的设计特点及多相机组合成像带来的增益,对比了不同方法的优劣。

3.1 基于传统阵列相机的宽视场高清成像方法

宽视场高清成像的探究从最初的超大胶片成像、单相机扫描拼接到基于阵列相机的复杂系统逐步发展成熟。Wilburn等^[27]设计了一个由100个定焦定光圈相机构成的阵列系统,系统中的子相机紧密排列,以近似单中心投影(single center of projection)。该系统由4台主机控制,利用现场可编程门阵列(FPGA)编程实现相机同步;每个子相

机自带一个计算单元,分担数据处理压力。子相机捕获的原始数据被压缩后传输给主机,以减轻数据传输压力,随后通过图像拼接技术进行配准融合,最终合成千万像素级宽视场高清视频。系统结构及对应的生成结果见图 4。由于系统中子相机间的成像

特性差异会影响配准的准确性,因此需要提前进行色彩校准、相机响应曲线估计等操作。后续一些工作对多相机系统中影响图像合成步骤的因素进行了相关分析研究,如 Ding 等^[28]对相机阵列的色彩校准研究。

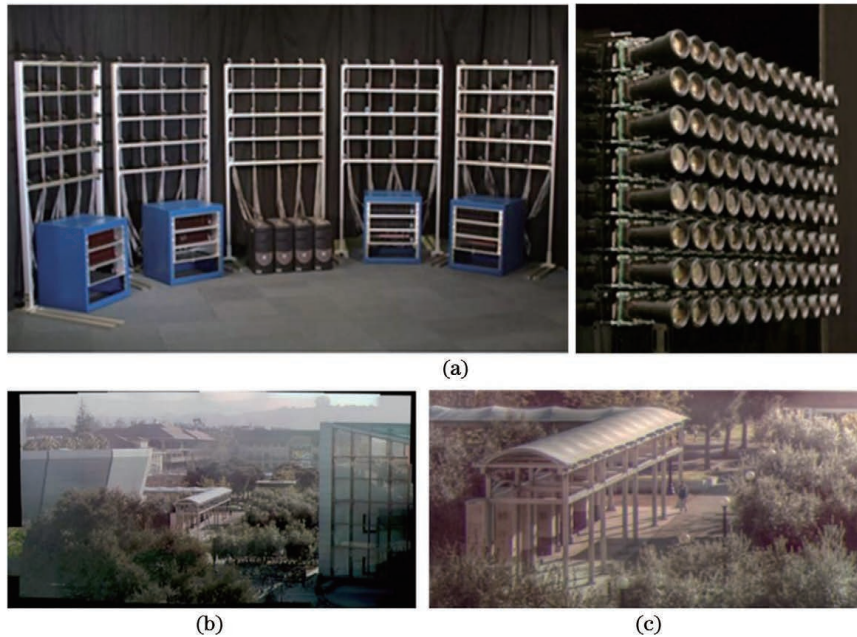


图 4 一种宽视场高清成像相机阵列及其合成结果^[27]。(a)相机阵列结构图;(b)由该阵列相机拍摄得到的宽视场超高清图像;(c)图(b)的局部放大视野

Fig. 4 A wide-field high-resolution imaging array camera and its synthesized image^[27]. (a) Architecture of the array camera; (b) captured image using the array camera; (c) zoomed-in view of (b)

Nomura 等^[12]设计了一种可弯曲的小型相机阵列(图 5 展示了 20 个子相机构成的相机阵列)。在拍摄过程中,用户可以用各种方式弯曲相机阵列来改变目标场景的采集结构(不会损坏系统),使得场

景内容的捕获更为灵活。他们利用 SIFT^[29](scale-invariant feature transform)特征点匹配和基于图的优化方法使拼贴碎片最小化,能够生成 15 frame/s 的拼贴视频(如图 5 下侧及右侧)。

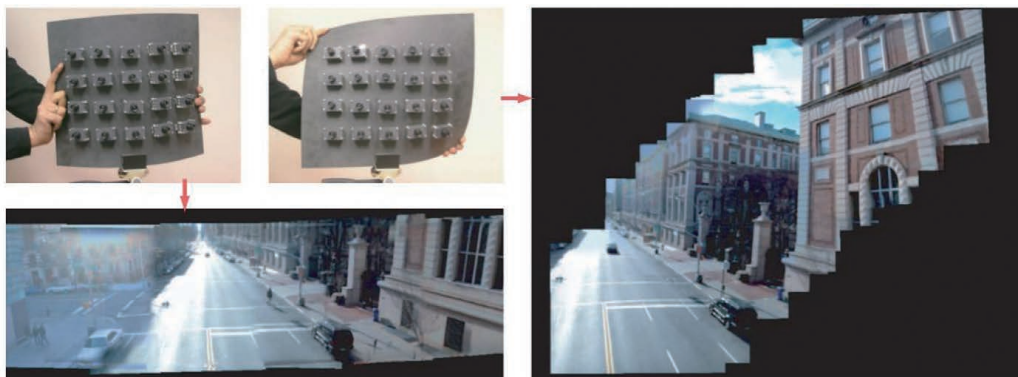


图 5 可弯曲相机阵列及其拍摄的街景图^[12]

Fig. 5 Flexible array camera and captured street scene image^[12]

3.2 基于多尺度阵列相机的宽视场高清成像方法

基于传统阵列相机设计的宽视场高清成像系统往往体积庞大,多尺度设计的提出实现了阵列相机

系统的小型化,成为宽视场超高清成像的主流方法。Brady 等^[25]发现基于微相机的成像系统的镜头速率和视场可以与镜头尺寸无关,继而基于此设计了基

于并行微型相机阵列的 AWARE-2 相机。AWARE-2 为单中心多尺度多孔径设计,由一个球对称的物镜和 98 个包围物镜的次级微相机组成(见图 6),能够以 3 frame/s 的速率拍摄 120° 视场 (field of view) 的十亿像素级图像。多尺度设计利用了现成相机阵列的现成焦平面 (off-the-shelf focal plane) 的优点,同时避免了多物镜成本和体积的缺

点。AWARE-2 实现了相机阵列的自适应测光,相比前述相机^[24,27]能够更大限度地利用子相机的动态范围。此外,Cossairt 等^[30]发现对于具有球面像差的透镜,去模糊误差并不随透镜尺度的增加而线性增加。他们基于此发现提出了比例定律 (scaling law),并设计了由一个球透镜和自带去模糊的传感器阵列构成的十亿像素级相机。

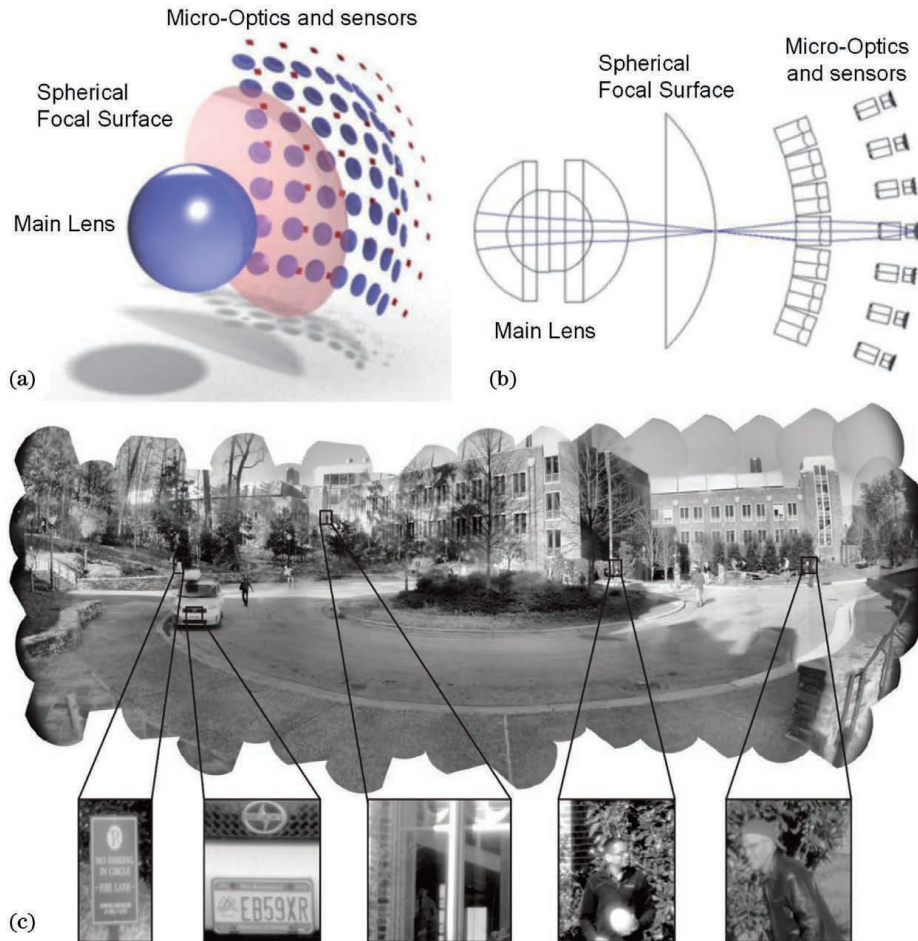


图 6 AWARE-2 的光学结构及重建结果。(a)(b)AWARE-2 的并行微型相机结构图^[25]; (c)AWARE-2 拍摄的宽视场超高清图像及其局部放大^[31]

Fig. 6 Optical structure and image reconstruction of AWARE-2. (a)(b) Structure of AWARE-2's parallel array of micro-camera^[25]; (c) an image captured using AWARE-2, of which the insets are digitally magnified images^[31]

上述相机虽然可以拍摄亿级宽视场高清内容,但需要复杂、费时的标定,且要求子相机间有一定的重叠区域以便图像匹配,灵活性和实用性受限。针对以上限制,Yuan 等^[32]搭建了一个分层模块化设计的非结构化阵列相机系统^[26],见图 7。该系统沿着模块内和模块间的路径协同工作,能够从捕获的多尺度视频序列中高效渲染实时 360° 视场的 3D 十亿像素级宽视场视频。同时,该系统能够根据时间熵 (temporal entropy map) 自适应地分配图像资源,降低了硬件成本,具有很高的灵活性、鲁棒性和

通用性。针对多尺度结构中主相机和子相机间视差、场景表现和色彩不一致导致图像匹配不准确的问题,Yuan 等^[33]提出了一种无需相机校准的具有一定鲁棒性的图像匹配和对齐算法,并解除了传统阵列相机校准过程要求子相机间有较大重叠区域的约束,为宽视场高清成像系统设计引入了更多的可扩展性和实用性。

多尺度多孔径成像方法目前已有相对成熟的相机系统应用,然而对于其应用和量产还有许多值得探索的方向,如系统功耗优化^[1]。

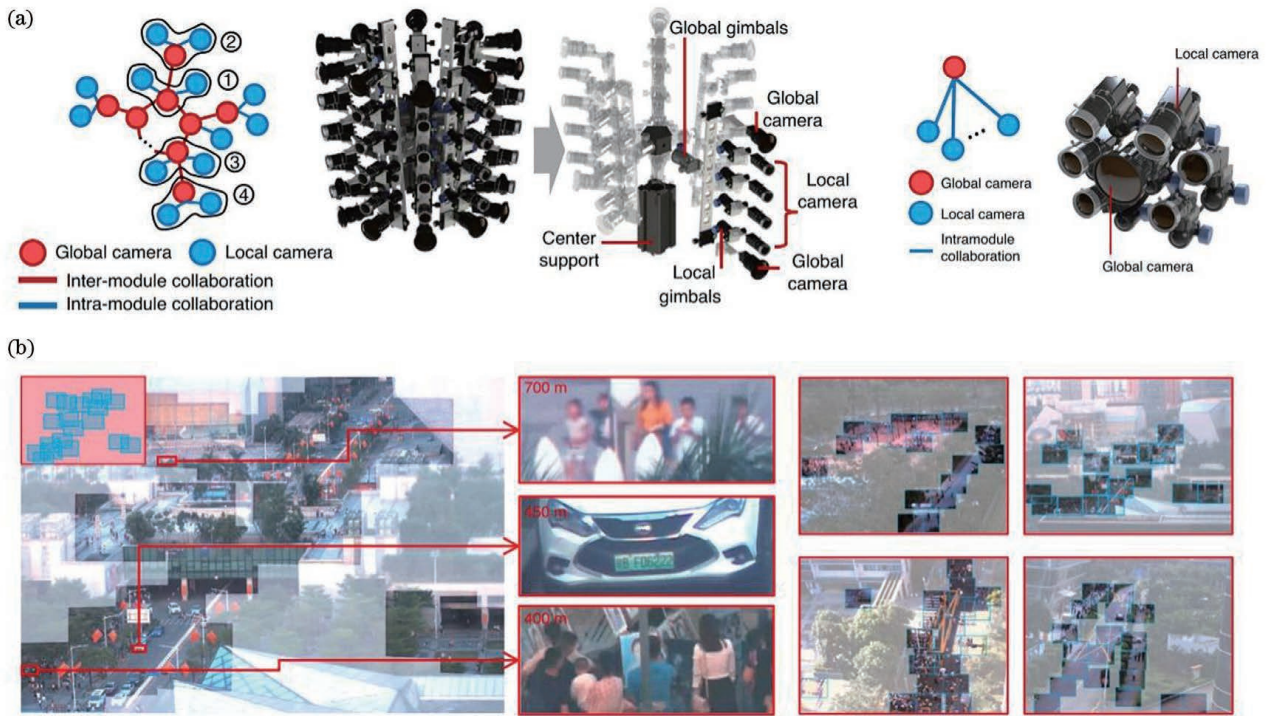


图 7 非结构化阵列相机^[26]。(a)相机系统及子模块结构图,相机和镜头的性质根据目标性质非均匀选择;(b)使用非结构化相机捕获的十亿像素级视频帧,红色和蓝色框分别表示全局和局部摄影机的分布

Fig. 7 The UnstructuredCam^[26]. (a) Schematic of UnstructuredCam system consisting of multiple columns of subarrays and UnstructuredCam module, and the cameras and lenses are selected heterogeneously based on the nature of the targets; (b) gigapixel-level videography captured using the UnstructuredCam, where the red and blue frames represent the distributions of the global and local cameras

4 时空高速视频采集

在当下消费者级别的成像市场中,相机的主流帧率一般在 30~90 frame/s 之间。但在燃烧研究、细胞成像、体育赛事分析等特殊应用场景中,高速运动过程中细微的时/空信息变化需要上百乃至上万帧率的高速成像设备才能准确记录,用以后续的研究分析。传统的高速摄像要求感光芯片的灵敏度大幅提升以及电路噪声强度进一步降

低,且对数据转化和存储速度有较高要求,这势必会提升成像模块的成本并限制了它的普及应用。此外,消费者追求更高清、更高速的内容获取,对便携型成像模组的需求日益增长。对单个普通相机拍摄的视频进行时间维度上的插值,很容易引起运动混叠(如图 8 所示),不能恢复真实的高速运动细节。因此,利用多相机系统实现时空高速(或高速高清)成像成为目前移动端高速成像的热门。

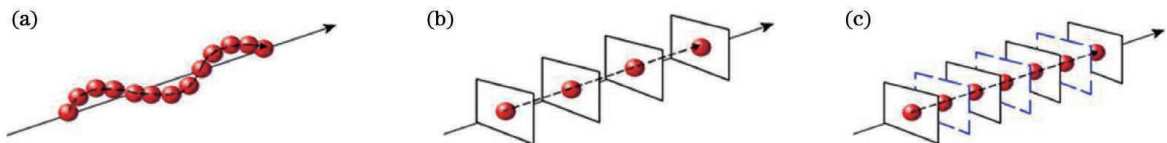


图 8 运动混叠现象^[34]。(a)红球以正弦状的轨迹在运动;(b)相机以较低帧率对该运动状态进行捕获所获得的帧序列,该帧序列中显示红球沿直线运动,这种错误的感知现象被称为运动混叠现象;(c)利用已有序列在时间维度上理想插值获得的帧序列(蓝色虚线矩形代表预测帧)无法恢复正确的运动信息

Fig. 8 Motion aliasing^[34]. (a) A ball moving in a sinusoidal trajectory; (b) an image sequence of the ball is captured at low frame-rate, the perceived motion is along a straight line, this false perception is referred to as “motion aliasing”; (c) the filled-in frames (indicated with blue dashed lines) which are obtained via an ideal temporal interpolation can not produce the correct motion

2002 年, Shechtman 等^[34]第一次提出利用多个低分辨率、低帧率相机输入恢复高帧率视频输出的方法, 他们利用不同的相机在不同时刻以相同的帧率采样, 从而恢复出了场景在整条时间线上的运动信息。2004 年, Wilburn 等^[35]进一步提出了利用多相机电子卷帘快门相继拍摄合成高清高速视频内容的思路。以上研究主要讨论了如何实现时间维度上更高细粒度的恢复。对于高速成像中的高清图象的获取, Wang 等^[36]、Cheng 等^[37]提出了用高分辨率、低帧率相机和低分辨率、高帧率相机同步触发的方式, 分路获取高分辨率空间信息或运动信息的方法, 实现了高分辨率空间信息与运动信息的融合, 进而实现了高速高清成像。

本章根据不同相机的工作协同方式, 将现有的基于传统相机的高速高清成像系统分为异步触发和同步触发两种类型进行介绍。

4.1 相同帧率的异步触发相机组合

异步触发组合是指不同的相机在不同的时刻开始记录运动信息, 通过不同帧序列在时间轴上的混合再采样获得整个时间轴上的运动信息。

2002 年, Shechtman 等^[34]首次提出了时空超分辨率的概念, 并针对多个时刻采集的低分辨率视频帧序列进行了高帧率内容的重建, 如图 9 所示。具体来说, 该方案对曝光时长引起的时间维度模糊以及相机点扩散函数引起的空间维度模糊进行联合建模。用 $p^l = (x^l, y^l, t^l)$ 表征低分辨率时空内的点, 用 $p^h = (x^h, y^h, t^h)$ 表征相应的高分辨率时空内的点, 两个域内点的对应关系为 $p^h = T_i(p_i^l)$ 。以 $B_i = B_{(\sigma_i, \tau_i, p_i^l)}$ 表示第 i 个低分辨率相机对点 p_i^l 进行的时空联合模糊函数, 那么理想的高分辨率时空域特征 $S(p^h)$ 与各个采样获得的低分辨率时空域特征 $S_i^l(p_i^l)$ 之间的关系可以表示为

$$S_i^l(p_i^l) = (S * B_i^h)(p^h) = \int_{x,y,t} S(p) B_i^h(p - p^h) dp. \quad (1)$$

$p = (x, y, t) \in \text{Support}(B_i^h)$

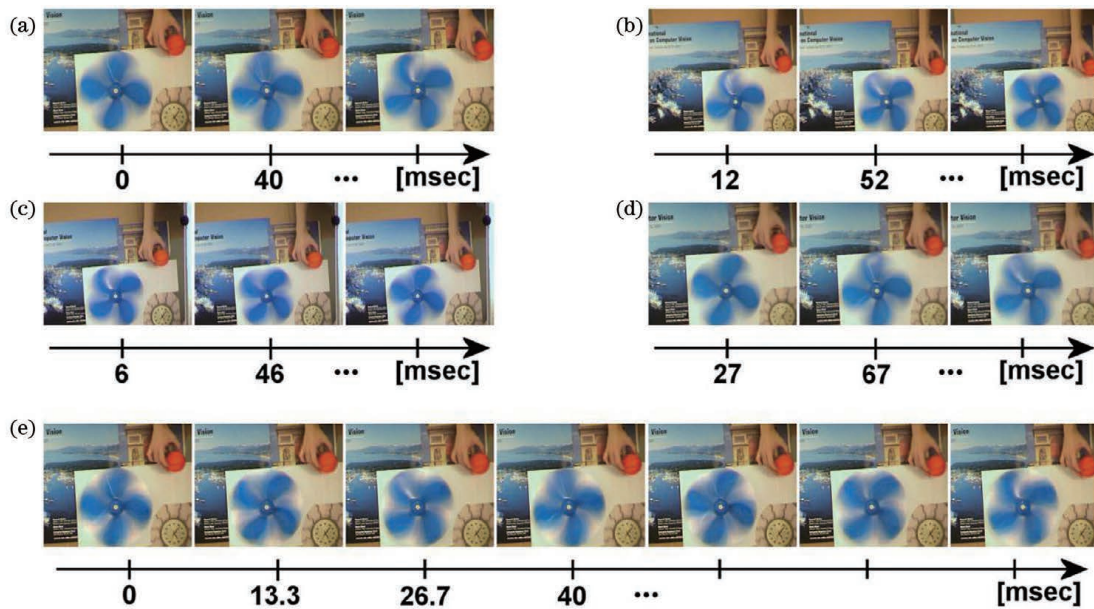


图 9 同帧率的异步触发相机组合^[34]。(a)~(d)不同相机在不同时刻开始拍摄的帧序列;(e)恢复的高帧率帧序列
 Fig. 9 Illustration of multi-cameras triggered with the same frame rate^[34]. (a)~(d)Captured sequences from different cameras at different time; (e) reconstructed sequences with high frame rate

通过多个采样的低分辨率时空域帧序列建立线性方程组, 提取其与高帧率视频之间的逆映射函数, 采用最小二乘法求解得到高帧率时空域信息, 从而实现均匀高帧率内容的重建过程。

Wilburn 等^[35]在此基础上利用密集普通相机阵列的卷帘快门相继拍摄生成了帧率为 1560 frame/s 的高速视频。如图 10 所示, 该方案通过控制相邻曝

光帧之间的空间位置相邻, 减少了帧间运动的不连续性, 同时, 在不同帧内提取相同时刻点曝光的数据行切片进行叠加形成最终的图片, 有效避免了卷帘快门带来的运动失真。相对于全局快门来说, 该方案可以实现更高帧率视频内容的获取与实时传输。

4.2 不同帧率的同步触发相机组合

同步触发组合是指不同的相机在同一时刻开始

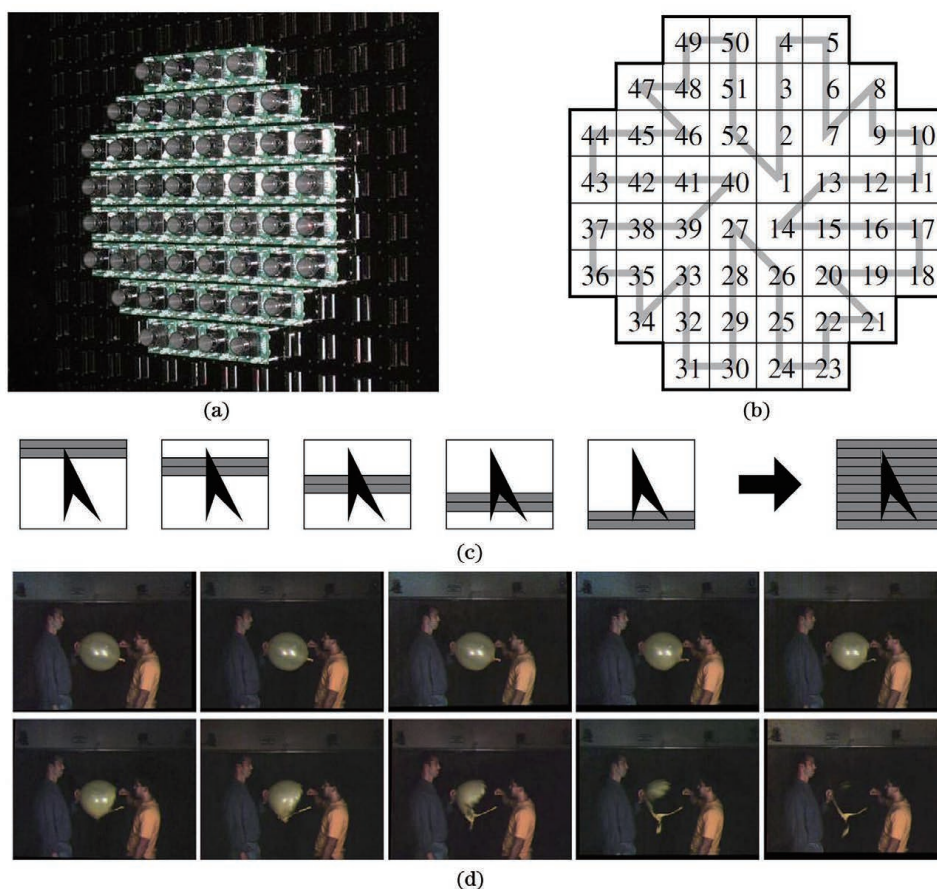


图 10 Wilburn 等的相机阵列示意图^[35]。(a)相机系统;(b)相机阵列的触发顺序;(c)从不同帧中获取相同时刻的数据切片生成最终图片,抑制电子卷帘快门引起的失真;(d)重建的气球被击破的高速视频片段

Fig. 10 Illustration of Wilburn *et al.*'s multiple camera array^[35]. (a) Detailed structure; (b) firing order for camera array; (c) the final view consists of different rows (shown in gray) in each camera at the same time, which suppresses the distortion caused by rolling shutter; (d) reconstructed high-speed video of a popping balloon

记录运动,相机间帧率各不相同。一般来说,该组合中各相机间的空间分辨率不尽相同,这样可以保证时空采样特征的多样性,在考虑高帧率的同时实现了高分辨率的保留。目前,高速高清成像系统主要使用的是低帧率、高分辨率相机与高帧率、低分辨率

相机的混合。

考虑到一般光场相机每秒钟仅能存储 3 帧,庞大的光场数据量使得现有相机无法记录视场内完整的运动信息,Wang 等^[36]提出了一种光场相机和普通相机相结合的混合相机系统,如图 11 所示,该相

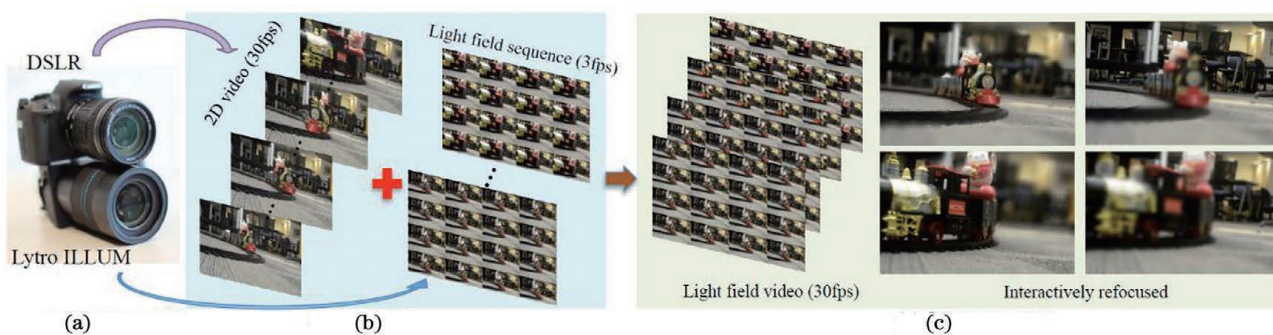


图 11 Wang 等所提方案的系统结构及输入、输出^[36]。(a)系统结构;(b)系统捕获的标准 30 frame/s 视频内容和 3 frame/s 光场序列;(c)最终生成的 30 frame/s 光场视频

Fig. 11 System setup and input/output proposed by Wang *et al.*^[36]. (a) Camera setup; (b) the inputs consist of a standard 30 frame/s video and a 3 frame/s light field sequence; (c) finally generated 30 frame/s light field video

机系统可以捕获 30 frame/s 的全光场视频内容。如图 12 所示,为了恢复光场相机在 t 时刻的输出,该方案分别计算出目标时刻的前后相同时刻光场相机捕获图像与标准相机捕获图像之间的视差关系,同时估算出标准相机图像从前后时刻变换至 t 时刻对应的运动光流,结合时刻间的相对关系,估计出了

t 时刻标准相机捕获图像与光场相机所得图像之间的视差图,并通过视差图获取了映射后的光场相机图像。考虑到估计视差、运动光流过程中对遮挡部分的处理会不可避免地带来误差,该方案进一步利用所有可参考的图像进行特征的深度融合,从而形成了最终的结果。

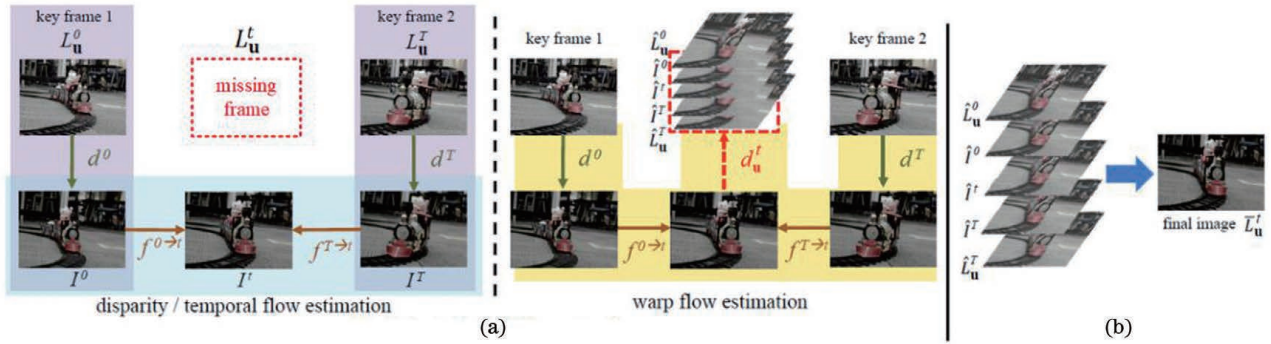


图 12 Wang 等提出的图像处理框架^[36]。(a) 估计关键帧间的视差以及视频帧在时间维度上的运动信息,从而预测目标帧时刻两路视角对应的视差图,实现从高帧率视频内容向低帧率视点的映射;(b) 融合所有参考图像形成最终的重建图像

Fig. 12 Overall image processing architecture proposed by Wang *et al.*^[36]. (a) Estimates the disparity at the key frames, as well as the temporal flow in the 2D video to generate the disparity at the target frame and then warps high frame-rate video to the low frame-rate viewpoint; (b) fuses all the reference images to output the final reconstruction

以上方案利用已知部分时刻的光场相机与标准相机对应帧之间的视差关系预测出每个时刻两者之间的视差图,从而实现标准相机空间特征向光场相机的转化,最后实现光场相机在时间轴上完整的运动信息恢复,获得高速高清成像。

低帧率的空间特征向低分辨率、高帧率视角转换,借助时间维度上的运动向量,将高分辨率信息传递到时间轴上的每个时刻。

Cheng 等^[37]从另一个角度出发,将高分辨率、

如图 13(b)所示,以高、低分辨率相机同时曝光的时刻 t_0 作为初始状态,计算两路视角在高分辨率空间上的相关性,从而获得真实高分辨率 I_{ref} 映射

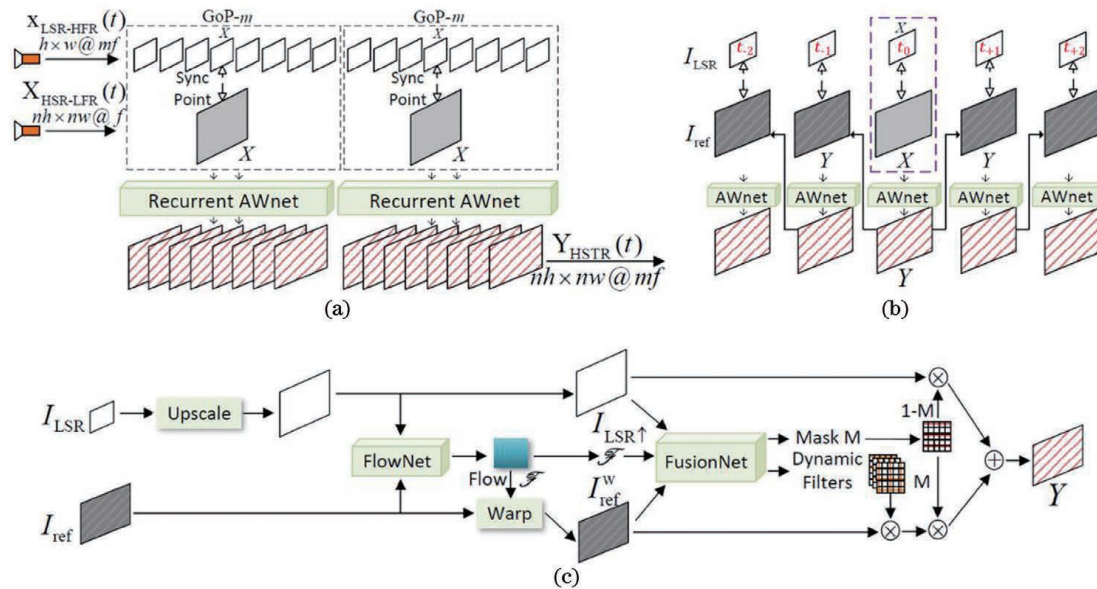


图 13 Cheng 等方案的具体阐述^[37]。(a) 双相机结构;(b) 高分辨率信息在时间维度上的传递过程;(c) 高、低分辨率融合的具体流程

Fig. 13 Illustration of Cheng *et al.*'s proposal^[37]. (a) Dual camera setup; (b) propagation of high-resolution map along with the temporal dimension; (c) the detailed pipeline of fusing the high resolution map and the low resolution map

到低分辨率视角的伪高分辨率特征 I_{ref}^w , 然后结合低分辨率视角上采样的空间信息 $I_{LSR\uparrow}$, 形成低分辨率相机所在视点在 t_0 时刻的高分辨率表征 Y 。之后, 以此时间点为中心向前后时刻 $(t_{-2}, t_{-1}, t_{+1},$

$t_{+2}, \dots)$ 进行逐帧高分辨率表征的传播, 确保高分辨率信息和低分辨率信息在时间维度上的相邻性, 最后实现低分辨率、高帧率视频内容在所有时刻的高分辨率重建, 实现高速高清成像(效果见图 14)。

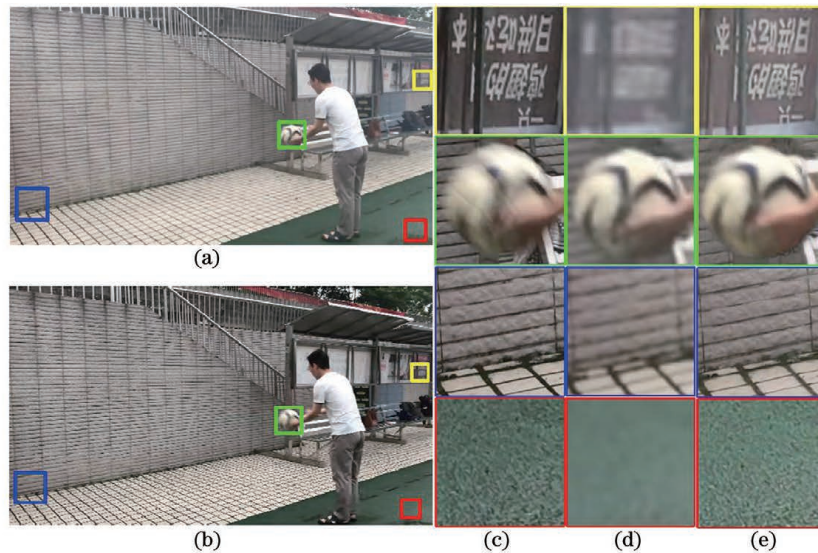


图 14 Cheng 等方案的可视化效果^[37]。(a) 高分辨率、低帧率的输入图 I_{ref} ; (b) 上采样的低分辨率、高帧率输入图 $I_{LSR\uparrow}$; (c)(d) I_{ref} 和 $I_{LSR\uparrow}$ 中对应的局部图片特写; (e) 该方案的重建图

Fig. 14 Visualized results of Cheng *et al.*'s proposal^[37]. (a) Input frame (I_{ref}) with high resolution and low frame rate; (b) input frame ($I_{LSR\uparrow}$) with low resolution and high frame rate; (c)(d) close-up of patches in I_{ref} and $I_{LSR\uparrow}$; (e) reconstructed frame with the proposal

当下基于传统相机的时空高速高清视频成像虽然在一定程度上突破了光场相机或普通高清相机在帧率上的限制, 但离业界对高速摄像的帧率要求仍有一定距离, 如何解决现有高清相机或光场相机在时间维度上更高密度的采样重建是后续有待解决的问题。

5 多相机高动态范围成像

传统相机难以捕获自然场景中的全部信息, 且

往往会引入噪声或信息丢失(传统相机成像的流程见图 15)。高动态(high dynamic range)成像旨在扩展输入数据的动态范围, 还原接近真实世界的场景细节, 已被广泛应用到消费电子、监控摄像、游戏渲染等领域。通过专用的高动态相机可以直接拍摄得到高动态图像, 但是这类专门设计的硬件设备需要配备特定的光学系统或图像传感器, 价格昂贵且校准流程复杂, 难以普及。

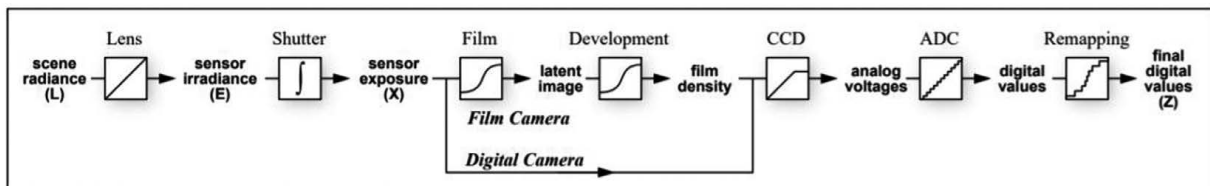


图 15 相机成像原理流程图^[38]

Fig. 15 Flow chart of camera imaging principle^[38]

随着图像处理技术的进步, 一些学者提出利用单张图像或多张不同曝光的图像来恢复不合理曝光区域的纹理细节(见图 16)。由于过曝程度较大时很难恢复图像饱和区域损失的信息, 因此基于多曝光图像的高动态重建目前仍是主流方法。多曝光图像序列的获取方式有两种: 一是利用多台相机, 且每

台相机设置不同的曝光参数; 二是利用单相机连续拍摄不同曝光的图像序列。基于第一种方式的重建相比第二种方式可以获得更宽的场景视野, 同时具有场景全部的时域信息, 但单相机多曝光方式每时刻的捕获帧中都存在信息丢失。此章节主要介绍第一种方式, 即基于多相机系统的高动态重建。

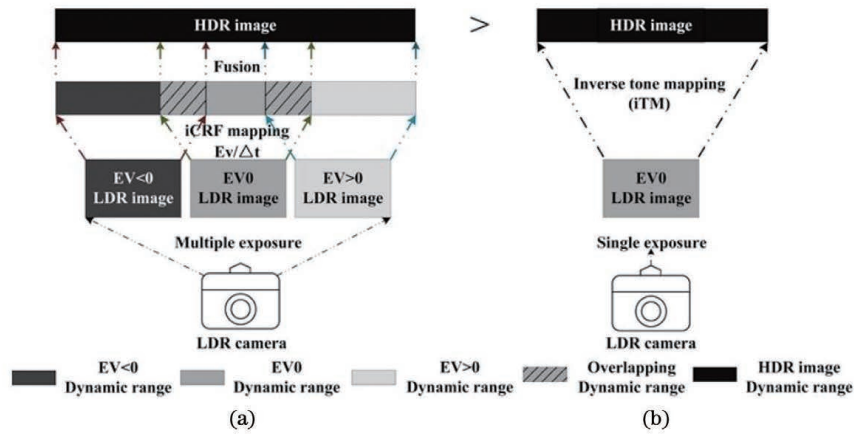


图 16 基于多曝光图像序列的高动态重建和基于反色调映射的高动态重建示意图^[39]。(a)基于多曝光图像序列的高动态重建;(b)基于反色调映射的高动态重建

Fig. 16 High dynamic range imaging based on multi-exposure image sequence and high dynamic range imaging based on inverse tone mapping^[39]. (a) High dynamic range imaging based on multi-exposure image sequence; (b) high dynamic range imaging based on inverse tone mapping

5.1 基于光学对齐的多相机高动态重建方法

基于硬件的多相机高动态重建方法一般利用分光镜辅助构建系统,分光镜两端为不同类型或功用的相机。Lee 等^[40]构建了由一台低分辨率相机、一台高分辨率相机和分光镜构成的混合相机系统,在该系统中,主相机接收分光镜的透射光拍摄高分辨率(high resolution)图像,辅相机接收反射光拍摄低分辨率(low resolution)图像。随后,利用遮挡区

域检测辅助光流计算,更准确地估计主相机和辅相机的帧间运动,进而重建高动态视频。Froehlich 等^[41]用两台相机分别捕获高光和暗光信息,两台相机的曝光差为 4 Fstop,通过一个 1 : 16 透过率的半透射镜使两台相机可以同步拍摄场景内容。该设计在相机使用长焦段时可能会导致伪影,因此部分情况下需要使用 50%透过率的分光镜配合减光滤光镜来保证拍摄的曝光比,如图 17 所示。

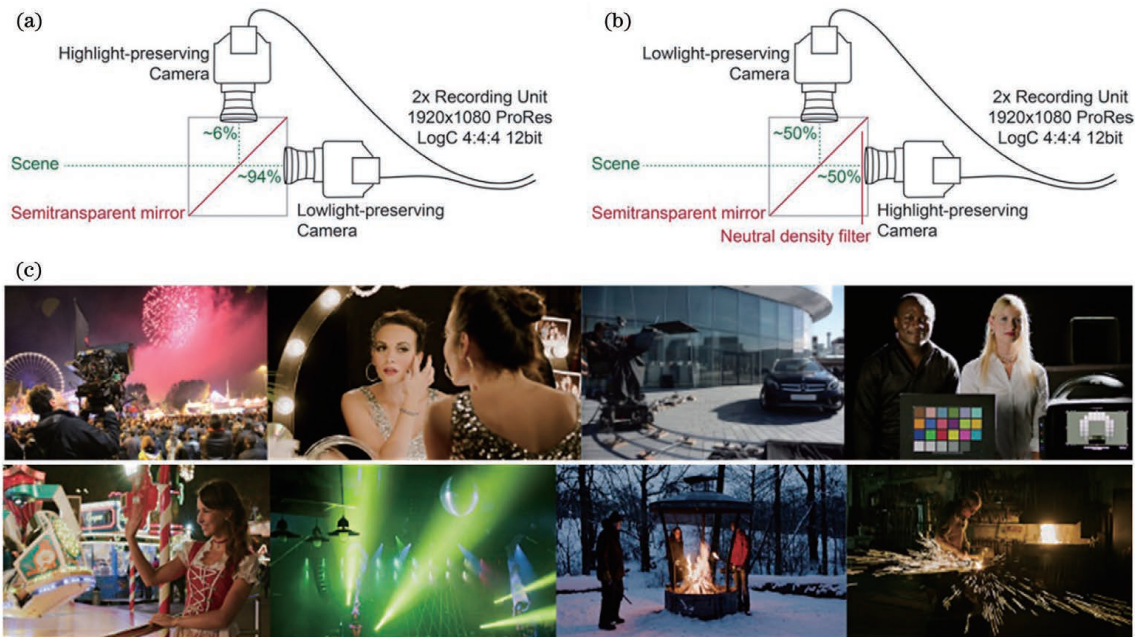


图 17 基于分光镜的高动态相机示意图及重建结果^[41]。(a)1 : 16 透过率的分光镜同步拍摄;(b)50%透过率的分光镜配合减光滤光镜拍摄;(c)高动态相机重建的场景

Fig. 17 Schematic and reconstructed frames of high dynamic range camera based on beam splitter^[41]. (a) Capture with 1 : 16 beam splitter; (b) capture with 50% semitransparent mirror and camera mounted with a neutral density filter; (c) reconstructed scene using the high dynamic range camera

Tocci 等^[42]受多传感器分色相机的启发设计了一个可以捕获高达 7 Fstop 动态范围的三传感器单镜头相机,如图 18 所示。该系统的三个传感器分别捕获高光、中间调和暗光信息,通过重复使用光路来提高光线利用率至 99.96%。该系统利

用两个分光镜使来自相机镜头的 92% 的光通过第一个分束器传输到 HE 传感器,7.52% 的光通过第一个分束器传输到 ME 传感器,0.44% 的光通过第一个分束器传输到 LE 传感器,动态范围扩展到约 7.7 Fstop。

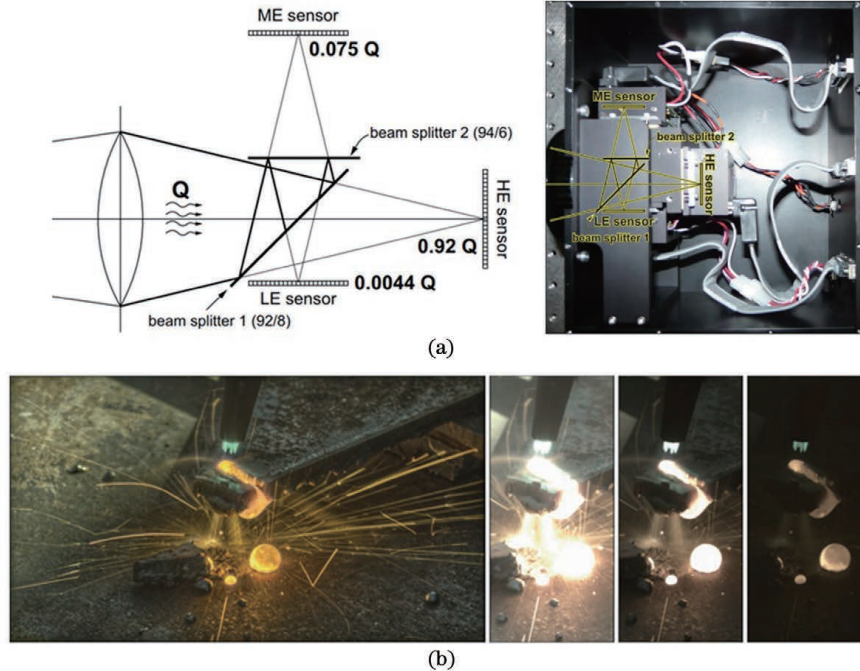


图 18 多传感器高动态相机及一组拍摄结果^[42]。(a)相机的光学结构,其中 LE、ME、HE 分别代表传感器接收的光的量为低、中、高;(b)最终的重建图像以及由 HE、ME、LE 传感器得到的低动态范围(LDR)图像

Fig. 18 Multi-sensor high dynamic range camera and a group of captured images^[42]. (a) Optical architecture of the high dynamic range camera, the terms high, medium, and low exposure (HE, ME, LE, respectively) refer to the sensors based on the amount of light each sensor receives; (b) final reconstructed image, where the inset photos show the low dynamic range images from the high, medium, and low-exposure sensors, respectively

5.2 基于图像处理的多相机高动态重建

已有的基于单帧图像增强进行高动态重建的方法一般采用逆色调映射^[43]实现动态范围的扩展^[44],如图 16(b)所示。由于输入图像的曝光范围有限,因此该方法在欠曝或过曝区域会存在过度增强引起的色彩失真或难以恢复的局部细节。基于单相机多曝光序列的高动态重建方法虽然记录了不同曝光条件下的场景信息,但在每时刻的捕获帧中都存在由遮挡等带来的信息丢失。

利用多相机系统拍摄多曝光序列,在捕获更广曝光范围的同时,更完整地记录了场景中空域和时域的有效信息,对重建高动态图像或视频提供了极大的增益。多曝光图像由不同的相机拍摄,因此需要对输入序列间的视差或运动进行补偿,以避免伪影。一般使用深度估计来实现多相机间的匹配,但相机间曝光水平的不同会给深度

估计带来一定难度,如何准确估计多曝光多视角图像的深度图是重建过程中的难点。Lin 等^[45]最早提出了一套多相机高动态重建的完整流程,包括相机响应曲线(CRF)估计、匹配对齐、融合重建和鬼影去除步骤。他们将不同曝光的图像利用 CRF 转换为相同亮度的图像,随后使用置信传播(belief propagation)进行立体匹配,通过设置阈值来检测对应像素对中的误匹配,以去除鬼影。相机响应曲线为图 15 中表示 X 到 Z 转换过程的函数 $Z = CRF(X)$ 。后续的基于深度的工作^[46-48]大多数延续了 Lin 等^[45]的重建流程,且对流程内部步骤进行了创新改进。Selmanovic 等^[46]使用 HDR-LDR 图像对重建双目高动态视频,该方法利用一路的 HDR 信息结合深度估计和查找表(LUT)重建另一路的高动态内容。Park 等^[48]利用累积分布函数(CDF)获取特征点来估计 CRF,

在超像素分割的基础上估计 HDR 图像的深度,并加入了空洞补全、边缘保留和平滑等后处理,以获取更好的对齐结果。Heo 等^[49]设置多相机为交替长短曝光,并对曝光设置相同的图像来估计深度图。而噪声或过曝对深度估计有不同程度的影响,因此得到的深度图会在不同区域出现错误。他们通过融合邻近深度图进一步对初始得到的深度图进行增强,最终基于联合双边滤波生成了高动态图像。Chen 等^[39]用三个卷积神经网络(CNN)分别模拟重建流程中的曝光调整、空洞补全和融合重建步骤,并提出了数据集受限情况下的训练方法。

此外,块匹配、可变形卷积也被运用到多相机

高动态重建工作的运动或视差估计中^[50-51]。Orozco 等^[51]将极限约束加入到块匹配(patch match)中,通过邻近搜索和最大化双向相似度迭代优化寻找最佳匹配块。确定块匹配关系后可得到对齐的图像,进而融合重建高动态图像。Pu 等^[50]利用可变形卷积对齐多尺度图像特征,随后将对齐的特征输入融合网络中重建参考视角的高动态图像,重建网络流程见图 19(a)。利用基于特征的偏移和金字塔对齐,该网络可以成功地处理视差和复杂运动,并且,通过亚像素级细化可以提高对齐精度。图 19(b)为该方法在一组多相机多曝光输入上的重建结果,可见,该方法能够较好地处理输入序列中的大视差。

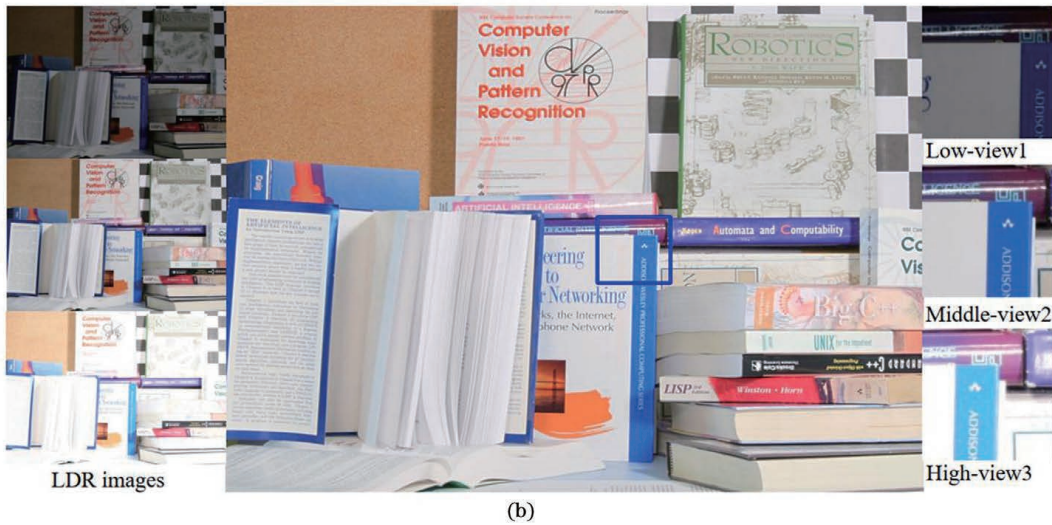
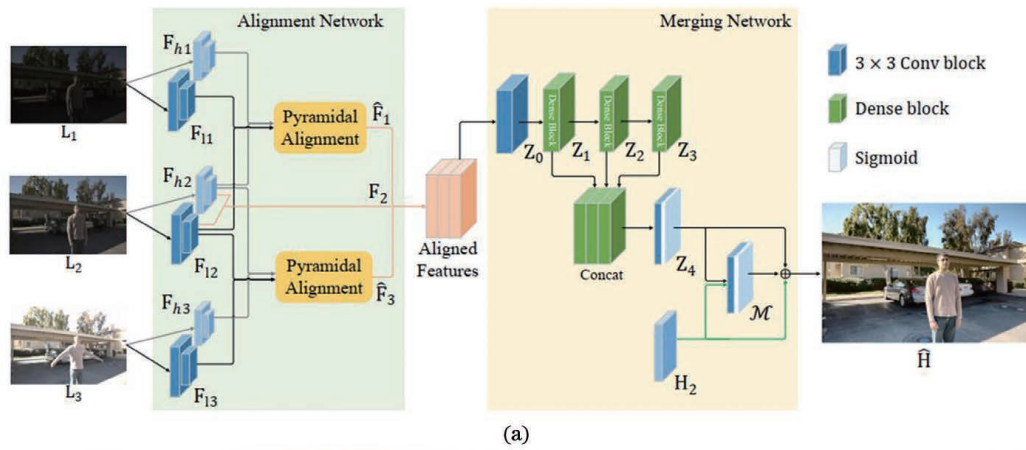


图 19 基于可变形卷积的高动态重建^[50]。(a)重建网络流程图;(b)多相机多曝光输入的高动态重建结果
Fig. 19 High dynamic range (HDR) reconstruction based on deformable convolution^[50]. (a) Flow chart of HDR reconstruction network; (b) HDR reconstruction on multi-exposure multi-camera input

高动态视频重建需要考虑时域一致性等多方面,同时相关数据集较少,因此目前已有的多相机高动态重建方法大多数仍然立足于多视角高动态图像重建,而多视角多曝光高动态视频重建的探索较少。

但随着 HDR10、杜比视界等相关标准及显示设备的升级更新及移动电子设备的多摄化趋势,多相机高动态视频将是未来一个重要的探索方向和应用需求。

6 低照度成像

当成像模块捕获的光子数量不够充分时(拍摄环境照度低、曝光时间过短或光圈较小),拍摄所得的图像往往表现出噪声强、可见度低、对比度低等特征,提供给用户视觉和机器视觉的有效信息大打折扣,严重影响了风光摄影、安防监控、自动驾驶等任务在低照度下的表现。延长曝光时间或增加外置闪光灯,虽然可以改善成像时的光照条件,但容易引起捕获图像中的运动模糊及不自然的眩目感。如何实现低照度条件下的优质成像一直是图像处理领域的热门问题。

早在 1990 年,Pizer 等^[52]就利用直方图均衡化来提升图像的全局亮度。而后基于 Retinex 理论的图像属性分解模型于 1997 年被应用于低光照图像的增强^[53],该模型通过对分离出来的反射图像和光照图像进行噪声抑制、亮度增强来推理理想曝光条件下该场景的成像。近些年来,由于深度学习技术的新兴发展,诸多研究^[54]利用深度神经网络在大规模数据集上提取低照度图像的非线性表达,实现了具有较好泛化性的端到端低照度图像的增强。以上方法主要面向单张低照度图像进行质量增强,结合成像的先验知识,全局调整图像亮度水平,优化噪声并改善色彩空间。但由于单图像输入的特征有限,重建后的图像在色彩表现和纹理细节方面缺乏一定的真实性。因此,可以利用多相机系统获取更加丰富的输入信息,为低照度条件下的可靠优质成像提供保障。

低照度条件下的多相机成像模块致力于利用成对的传感器在有限的时间内捕获更多的光子并将其

转化为电信号,模块中可使用多个成像参数(快门、分辨率、光圈等)不同的相同传感器或者成像参数相同的不同传感器来丰富低照度下的信号采集,且模块中也保留了一般条件下的彩色成像,可有效规避长曝光时间带来的模糊效应,保障最终成像结果真实性。以下主要介绍当下三种典型的相机组合形式,并对其性能进行分析。

6.1 有外置光源与无外置光源的相机组合

延长曝光时间、加大光圈或者提高相机增益等方法在改善低照度成像质量的同时会引起运动模糊、景深有限、噪声放大等弊端,而闪光成像可以提供外在瞬间光源,改善周围光照环境,保留高频纹理细节,提升图像的信噪比,且不需要长时间曝光、大光圈或者高相机增益。但闪光光源一般为点光源,会导致场景中的物体受光不均匀,且闪光光源与环境光的差异会破坏原有场景的氛围,同时也会带来红眼效应或者尖锐的阴影等。

Petschnigg 等^[55]采用无闪光拍摄与闪光拍摄相结合的方式进行了低照度条件下的成像,用噪声水平低的闪光成像辅助去噪,将闪光图像中的高频纹理细节补充到无闪光图像中,在提升图像信噪比的同时保留了无闪光条件下的环境光分布,使得最终的成像更接近于场景的真实分布。此外,闪光图像与无闪光图像中瞳孔的变化也可以帮助定位瞳孔位置,从而削弱闪光带来的红眼效应。该方案在正常成像的基础上融合了闪光成像,并且融合了周围环境光和外在强光光源条件下的场景特征,在提升无闪光拍摄图像质量的同时有效地抑制了外置闪光带来的负面效应,极大地改善了低照度条件下的成像水平(效果见图 20)。

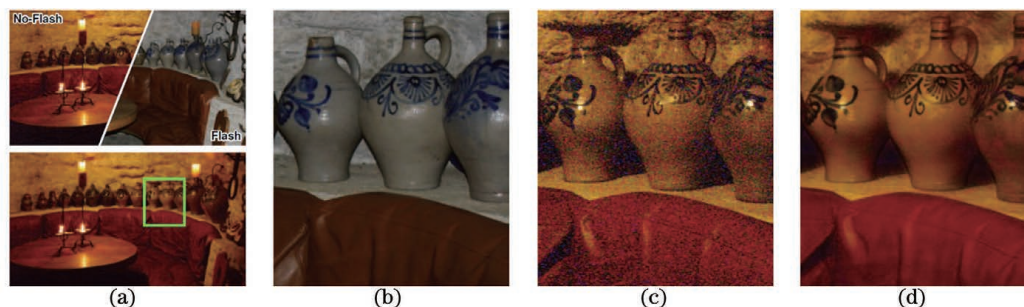


图 20 无外置光源成像与有外置光源成像的对比^[55]。(a) 原始无外置光源成像、有外置光源成像以及两者结合的成像结果;(b)~(d) 三种成像结果的局部细节效果图

Fig. 20 Illustration of imaging with flash and without flash^[55]. (a) Standard photography, flash photography and the combined photography;(b)-(d) enlarged views of the local details

6.2 可见光与不可见光的相机组合

尽管外置闪光灯给低照度条件下的成像提供了

足够的亮度,但成像时也会由于瞬间爆发光照而在平滑表面上形成不自然的高光,而且使被拍摄者有

眩目感。为了克服外置闪光灯带来的困扰, Krishnan 等^[56]提出了暗闪光技术。暗闪光是指去除了可见光波段的外置光源,其绝对辐照度如图 21(b)所示。由图 21 可知,相机对不同波段的光谱响应曲线各异,各通道像素值的幅度也有较大差别,但各波段成像在场景梯度特征上保留了极强的相关性。暗

闪光技术基于该强相关性通过外置光源增加了不可见光波段的瞬时成像光子数量,并通过保留场景在不可见光波段的梯度特征来增强可见光波段的场景信息,既不影响拍摄对象,也允许了更快的快门速度,抑制了相机抖动带来的动态模糊;但不可见光波段的成像难以提供与人眼视觉感受相匹配的色彩。

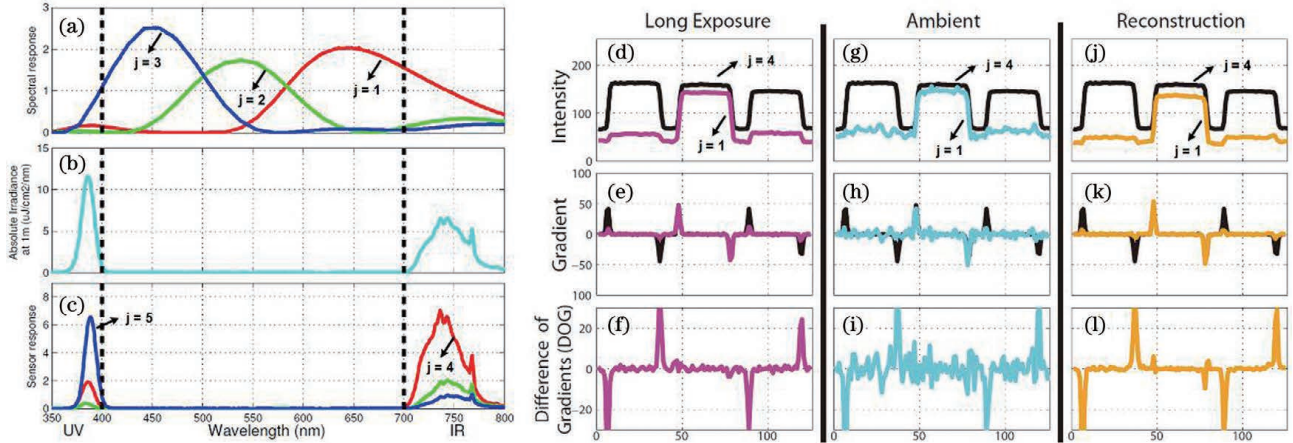


图 21 传统 RGB 成像与暗闪光成像的光谱响应曲线说明及不同通道对应亮度的对比^[56]。(a)(c)传统 RGB 成像和暗闪光成像的光谱响应曲线, $j \in \{1, 2, 3, 4, 5\}$ 分别对应可见光红色、绿色、蓝色以及红外、紫外波段; (b)暗闪光光源在 1 m 处对应的绝对辐照度; (d)~(l)低光照条件下,长曝光 RGB 成像、环境光 RGB 成像以暗闪光成像在相同位置处可见光红色和红外波段的亮度及其梯度对比

Fig. 21 Illustration of spectral response curves for traditional RGB imaging and dark flash imaging along with the luminance comparison between different channels^[56]. (a) (c) Spectral response curves for RGB imaging and dark flash imaging, $j \in \{1, 2, 3, 4, 5\}$ for red, green, blue, IR, and UV channels; (b) absolute irradiance at 1 m from the dark flash; (d)–(l) under low-light condition, luminance and its gradient comparison between the red channel and IR channel of the corresponding pixel in the long-exposure imaging, ambient light imaging, and dark flash imaging

在此基础上, Wang 等^[57]提出了采用多相机系统与暗闪光技术结合的方式来克服不可见光波段成像弊端的方案。该方案中,他们将标准 RGB 相机中的红蓝通道换成近红外 (NIR)、近紫外 (NUV) 通道构成 NIR-G-NUV 相机,同时搭配标准的 RGB 相机构成双摄系统,使用暗闪光技术实现了低照度

成像,见图 22。相较于先前的暗闪光成像技术中仅使用不可见光波段在环境光和闪光条件下的相继成像,该方案中 RGB 相机和 NIR-G-NUV 相机同时拍摄,规避了分时拍摄带来的动态模糊。该方案中的标准 RGB 模组在低照度下成像虽然有比较强的噪声,但保留了与人眼感知信息相匹配的 RGB

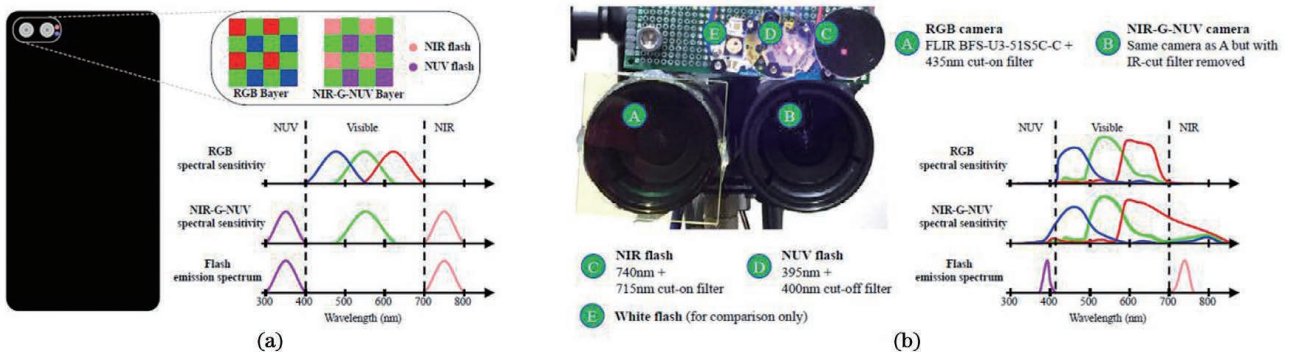


图 22 可见光与不可见光波段成像组合示意图^[57]。(a)理想相机原型; (b)实际搭建系统及光谱曲线

Fig. 22 Illustration of hybrid imaging setup with RGB and NIR-G-NUV cameras^[57]. (a) An idealized prototype; (b) an actual camera system and spectral curves

信息,利用暗闪光成像的图片可对其进行质量增强并抑制噪声。此外,该方案中的双摄模组均有绿色波段成像,利用此进行视角间的特征匹配可有效降低不同源图像对齐的难度。

该方案考虑到了可见外置闪光光源会对成像对

象产生的干扰,结合传统 RGB 成像与暗闪光成像,利用可见光波段标准成像与不可见光波段闪光成像同时在低照度条件下拍摄,以闪光成像为参考恢复理想 RGB 成像效果(见图 23),抑制了噪声,实现了真实可信的低噪高质量彩色图像。

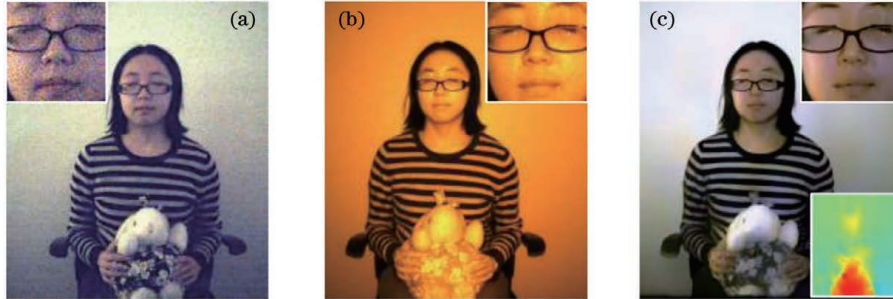


图 23 Wang 等方案的效果图^[57]。(a)低照度 RGB 相机输入(这里显示了 5 倍数字增益后的图像,便于观察);(b)暗闪光相机输入;(c)输出

Fig. 23 Visualization of Wang *et al.*'s method^[57]. (a) RGB input (visualized here with a digital gain of 5×); (b) dark-flash input; (c) output

6.3 单通道与多通道相机组合

上文中提及的相机模组使用不可见光波段在低照度条件下进行光电转换,虽然降低了强可见光对人眼造成的眩晕感,但传感器前的滤波片仍造成了部分光信号损失,且容易带来色彩空间的失真。为了最大程度地保留入射光信号,Dong 等^[58]和 Guo 等^[59]学者提出了使用单通道相机和彩色相机结合的组合方式来实现低照度成像的方法。单通道相机中舍弃了传感器前端的彩色滤波片,虽然无法恢复相应的色彩信息,但凭借其相对更高的光通过率,可以更细腻地保存场景的结构变化,保证了低照度下的清晰成像。

2015 年 8 月,奇酷 360 首次推出单通道与多通道结合(monochrome+RGB)的旗舰版双摄手机,

该手机利用单通道成像为彩色成像补充场景的高频纹理细节,因此在低照度条件下具有出色的成像结果。该双摄组合后来也成为主流的双摄配置,在华为、OPPO 等手机中加以应用^[60]。不同于以往利用单通道图像进行结构信息增强的思路,Dong 等^[58]利用彩色相机获取的色彩信息给细节高清的灰度图像进行上色,实现了高质量成像。他们利用输入图片对在局部灰度空间的相似性来寻找上色的最佳参考,因而对于那些被遮挡的部分,也可以借助局部的灰度分布找到一个最接近的上色区域来辅助上色,最终实现单通道图像的高质量成像。

在此基础上,Guo 等^[59]采用混合尺度的单通道相机与彩色相机组合系统进行低照度条件下的成像,成像流程如图 24 所示。受人眼对色度敏感性低

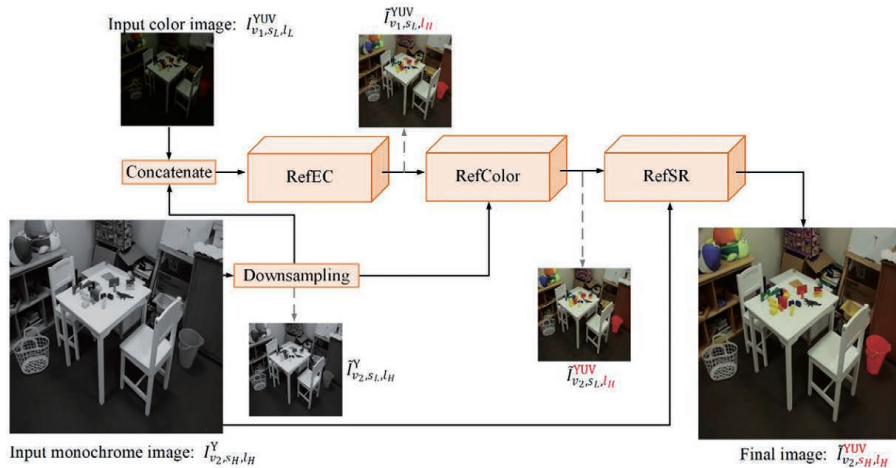


图 24 混合尺度的单通道相机与彩色相机实现的低照度高质量彩色成像流程图^[59]

Fig. 24 Workflow of low-light and high-quality color imaging with hybrid monochrome and color cameras^[59]

于亮度这一视觉行为的启发,Guo 等利用高分辨率单通道相机进行空间采样,利用低分辨率彩色相机进行色彩信息捕获,获得了高质量彩色图像重建所需的场景特征信息,而后借助数据驱动的级联神经网络模块实现了相机对间的混合尺度特征补偿及多维空间融合,最终获得了低照度下的高质量成像。

该方案对输入图像对中特征空间的差异性进行了有参考的补偿,逐步减小输入图像对之间的特征差距,实现了不同视角下的特征匹配迁移。具体来

说,该方案利用光信息丰富的单通道图像在低分辨率尺度上对彩色图像进行曝光调整(RefEC 模块),使彩色图像与单通道图像的曝光水平一致。接着,在 RefColor 模块中,利用增强后的彩色图像和单通道图像在亮度空间上的相似性计算两个视角间的对应像素的位移,从而实现了单通道图片在低分辨率尺度上的色彩恢复。最后,将得到的同一个视角的低分辨率彩色信息和高分辨率结构信息输入到 RefSR 模块中进行色彩空间的插值,最终获得了高分辨率的高质量彩色图像(结果见图 25)。

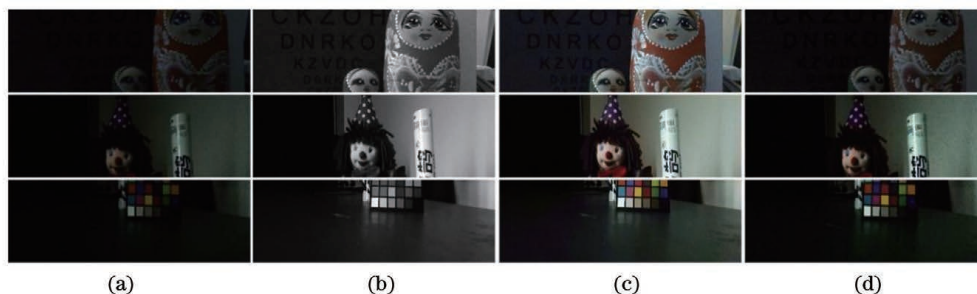


图 25 Guo 等的方案在低照度实拍数据集上的效果细节可视化^[59]。(a)低分辨率 RGB 输入图像(此处为了方便对比,将低分辨率输入放大到高分辨率大小);(b)高分辨率单通道输入图像;(c)重建后的高分辨率 RGB 图像;(d)对(a)图进行数字增益,用以展示噪声与模糊现象

Fig. 25 Close-up performance of Guo et al.'s method on the real scenes in the low-light condition^[59]. (a) Low-resolution RGB input images (for comparison, the low-resolution inputs are enlarged to the high-resolution sizes); (b) high-resolution monochromatic images; (c) reconstructed high-resolution RGB images; (d) the value-amplified low-resolution images from (a) for showing noise and blur

该方案将低照度成像问题化解为混合尺度单通道图像与彩色图像融合的问题,利用输入图像对在不同特征维度空间上的多样性信号进行分辨率空间、色彩空间的高分辨率重建,避免了长曝光引起的动态模糊,也保证了输出图像色彩的真实性。

多相机低照度成像系统架构繁多,成像形式丰富,重建内容更加稳定可靠。但当前的研究工作更多关注于静止场景或运动较小的场景的拍摄,对于低照度下运动信息的有效重建有待后续研究。

7 结束语

受生物双目和多目视觉系统启发的多相机成像系统,通过对不同维度、不同尺度的多源信息输入进行计算融合,实现了更高质量的场景重建,降低了高质量成像对单一相机模组光学硬件的要求。本文逐一介绍了多相机成像系统在宽视场超高清成像、时空高速视频采集、高动态范围成像及低照度成像四个方面的成像增强应用,以期帮助读者理解多相机成像系统在成像质量增强方面的优势。相对于单相机成像时的启发式增强,多相机系统利用不同的相

机分别保留了原始场景更多维度的高分辨率采样,而后基于空域/特征域的配准实现各相机间多维特征的参考融合,保障了全系统内各维度成像增强表现的真实性、可靠性,从而更加鲁棒地实现高质量场景重建。但如何利用多相机系统实现更优的低维耦合离散信号采样,提升高质量成像能力是需要进一步深究的问题。

本文主要关注多相机系统在成像视场范围、成像时空分辨率以及动态范围等方面的增强表现。对于多相机系统在显微成像、高光谱高清成像、遥感成像等领域的应用值得读者进一步关注。同时,文中提及的多相机系统在普及应用时如何做到轻量化、小型化还需要学术界和产业界的共同努力。更进一步,人工智能技术的飞速发展带领着计算成像技术实现新的突破,如何更深入地在多相机系统中将人工智能技术与光学系统设计联合起来也是非常值得研究的新兴方向^[10,61]。

参 考 文 献

- [1] Brady D J, Pang W B, Li H, et al. Parallel cameras

- [J]. *Optica*, 2018, 5(2): 127-137.
- [2] Dai Q H, Suo J L, Ji X Y, et al. Computational photography: computational capture of plenoptic visual information[M]. Beijing: Tsinghua University Press, 2016.
戴琼海, 索津莉, 季向阳, 等. 计算摄影学: 全光视觉信息的计算采集[M]. 北京: 清华大学出版社, 2016.
- [3] Barbastathis G, Ozcan A, Situ G H. On the use of deep learning for computational imaging[J]. *Optica*, 2019, 6(8): 921-943.
- [4] Cao X, Yue T, Lin X, et al. Computational snapshot multispectral cameras: toward dynamic capture of the spectral world[J]. *IEEE Signal Processing Magazine*, 2016, 33(5): 95-108.
- [5] Manekar R, Tayal K, Kumar V, et al. End-to-end learning for phase retrieval[C]//ICML workshop on ML Interpretability for Scientific Discovery, July 13-18, 2020, Virtual Event. Cambridge: PMLR, 2020.
- [6] Yao Y, Zuo M, Dong Y, et al. Polarization imaging feature characterization of different endometrium phases by machine learning [J]. *OSA Continuum*, 2021, 4(6): 1776-1791.
- [7] Suo J L, Liu Y B, Ji X Y, et al. Computational photography: keys, methods and applications [J]. *Acta Automatica Sinica*, 2015, 41(4): 669-685.
索津莉, 刘焯斌, 季向阳, 等. 计算摄影学: 核心、方法与应用[J]. *自动化学报*, 2015, 41(4): 669-685.
- [8] Adelson E H, Bergen J R. The plenoptic function and the elements of early vision [M] // Landy M, Movshon J A. *Computational models of visual processing*. Cambridge: The MIT Press, 1991.
- [9] Zuo C, Feng S J, Zhang X Y, et al. Deep learning based computational imaging: status, challenges, and future [J]. *Acta Optica Sinica*, 2020, 40 (1): 0111003.
左超, 冯世杰, 张翔宇, 等. 深度学习下的计算成像: 现状、挑战与未来[J]. *光学学报*, 2020, 40(1): 0111003.
- [10] Shao X P, Liu F, Li W, et al. Latest progress in computational imaging technology and application [J]. *Laser & Optoelectronics Progress*, 2020, 57 (2): 020001.
邵晓鹏, 刘飞, 李伟, 等. 计算成像技术及应用最新进展 [J]. *激光与光电子学进展*, 2020, 57 (2): 020001.
- [11] Fang L, Dai Q H. Computational light field imaging [J]. *Acta Optica Sinica*, 2020, 40(1): 0111001.
方璐, 戴琼海. 计算光场成像 [J]. *光学学报*, 2020, 40(1): 0111001.
- [12] Nomura Y, Zhang L, Nayar S K. Scene collages and flexible camera arrays [C] // *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, May 25, 2007, Grenoble, France. Geneva: Eurographics Association, 2007: 127-138.
- [13] Lin X, Liu Y B, Wu J M, et al. Spatial-spectral encoded compressive hyperspectral imaging [J]. *ACM Transactions on Graphics*, 2014, 33(6): 1-11.
- [14] Land M F, Nilsson D E. *Lens eyes on land* [M] // *Animal eyes*. Oxford: Oxford University Press, 2012: 94-129.
- [15] Mildenhall B, Srinivasan P P, Tancik M, et al. NeRF: representing scenes as neural radiance fields for view synthesis [M] // Vedaldi A, Bischof H, Brox T, et al. *Computer vision-ECCV 2020. Lecture notes in computer science*. Cham: Springer, 2020: 405-421.
- [16] Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [17] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision [C] // *IJCAI '81: Proceedings of the 7th International Joint Conference on Artificial Intelligence*, August 24-28, 1981, Vancouver, BC, Canada. San Francisco: William Kaufmann, 1981.
- [18] de Tone D, Malisiewicz T, Rabinovich A. Deep image homography estimation [EB/OL]. (2016-06-13) [2021-05-25]. <https://arxiv.org/abs/1606.03798>.
- [19] Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks [C] // *2017 IEEE International Conference on Computer Vision (ICCV)*, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 764-773.
- [20] Liu Y, Chen X, Wang Z F, et al. Deep learning for pixel-level image fusion: recent advances and future prospects [J]. *Information Fusion*, 2018, 42: 158-173.
- [21] Kaur H, Koundal D, Kadyan V. Image fusion techniques: a survey [J]. *Archives of Computational Methods in Engineering*, 2021: 1-23.
- [22] Li S T, Kang X D, Hu J W. Image fusion with guided filtering [J]. *IEEE Transactions on Image Processing*, 2013, 22(7): 2864-2875.
- [23] Lohmann A W. Scaling laws for lens systems [J]. *Applied Optics*, 1989, 28(23): 4996-4998.
- [24] Kopf J, Uyttendaele M, Deussen O, et al. Capturing and viewing gigapixel images [J]. *ACM Transactions on Graphics*, 2007, 26(3): 93.
- [25] Brady D J, Gehm M E, Stack R A, et al. Multiscale gigapixel photography [J]. *Nature*, 2012, 486

- (7403): 386-389.
- [26] Zhang J N, Zhu T Y, Zhang A K, et al. Multiscale-VR: multiscale gigapixel 3D panoramic videography for virtual reality [C] // 2020 IEEE International Conference on Computational Photography (ICCP), April 24-26, 2020, St. Louis, MO, USA. New York: IEEE Press, 2020: 1-12.
- [27] Wilburn B, Joshi N, Vaish V, et al. High performance imaging using large camera arrays[J]. *ACM Transactions on Graphics*, 2005, 24(3): 765-776.
- [28] Ding C Q, Ma Z. Multi-camera color correction via hybrid histogram matching[EB/OL]. [2020-05-31]. <https://box.nju.edu.cn/f/ef598961b2bf4c1a8dae/>.
- [29] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [30] Cossairt O S, Miao D, Nayar S K. Gigapixel computational imaging[C]//2011 IEEE International Conference on Computational Photography (ICCP), April 8-10, 2011, Pittsburgh, PA, USA. New York: IEEE Press, 2011: 1-8.
- [31] Son H S, Marks D L, Hahn J, et al. Design of a spherical focal surface using close-packed relay optics [J]. *Optics Express*, 2011, 19(17): 16132-16138.
- [32] Yuan X Y, Ji M Q, Wu J M, et al. A modular hierarchical array camera [J]. *Light, Science & Applications*, 2021, 10(1): 37.
- [33] Yuan X Y, Fang L, Dai Q H, et al. Multiscale gigapixel video: a cross resolution image matching and warping approach[C]//2017 IEEE International Conference on Computational Photography (ICCP), May 12-14, 2017, Stanford, CA, USA. New York: IEEE Press, 2017: 1-9.
- [34] Shechtman E, Caspi Y, Irani M. Increasing space-time resolution in video[M]//Heyden A, Sparr G, Nielsen M, et al. *Computer vision-ECCV 2002. Lecture notes in computer science*. Heidelberg: Springer, 2002, 2350: 753-768.
- [35] Wilburn B, Joshi N, Vaish V, et al. High-speed videography using a dense camera array [C] // Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, June 27-July 2, 2004, Washington, DC, USA. New York: IEEE Press, 2004: II.
- [36] Wang T C, Zhu J Y, Kalantari N K, et al. Light field video capture using a learning-based hybrid imaging system[J]. *ACM Transactions on Graphics*, 2017, 36(4): 1-13.
- [37] Cheng M, Ma Z, Asif S, et al. A dual camera system for high spatiotemporal resolution video acquisition[J/OL]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. [2020-05-31]. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9052471>.
- [38] Debevec P E, Malik J. Recovering high dynamic range radiance maps from photographs [C] // ACM SIGGRAPH 2008 classes on-SIGGRAPH '08, August 11-15, 2008, Los Angeles, California. New York: ACM Press, 2008: 1-10.
- [39] Chen Y Y, Jiang G Y, Yu M, et al. Learning stereo high dynamic range imaging from a pair of cameras with different exposure parameters [J]. *IEEE Transactions on Computational Imaging*, 2020, 6: 1044-1058.
- [40] Lee B, Song B C. Multi-image high dynamic range algorithm using a hybrid camera [J]. *Signal Processing: Image Communication*, 2015, 30: 37-56.
- [41] Froehlich J, Grandinetti S, Eberhardt B, et al. Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays [J]. *Proceedings of SPIE*, 2014, 9023: 90230X.
- [42] Tocci M D, Kiser C, Tocci N, et al. A versatile HDR video production system[J]. *ACM Transactions on Graphics*, 2011, 30(4): 1-10.
- [43] Huo Y Q, Yang F, Dong L, et al. Physiological inverse tone mapping based on retina response[J]. *The Visual Computer*, 2014, 30(5): 507-517.
- [44] Banterle F, Ledda P, Debattista K, et al. Expanding low dynamic range videos for high dynamic range applications [C] // Proceedings of the 24th Spring Conference on Computer Graphics-SCCG '08, April 21-23, 2008, Budmerice, Slovakia. New York: ACM Press, 2010: 33-41.
- [45] Lin H Y, Chang W Z. High dynamic range imaging for stereoscopic scene representation[C]//2009 16th IEEE International Conference on Image Processing (ICIP), November 7-10, 2009, Cairo, Egypt. New York: IEEE Press, 2009: 4305-4308.
- [46] Selmanovic E, Debattista K, Bashford-Rogers T, et al. Enabling stereoscopic high dynamic range video [J]. *Signal Processing: Image Communication*, 2014, 29(2): 216-228.
- [47] Bätz M, Richter T, Garbas J U, et al. High dynamic range video reconstruction from a stereo camera setup [J]. *Signal Processing: Image Communication*, 2014, 29(2): 191-202.
- [48] Park W J, Ji S W, Kang S J, et al. Stereo vision-based high dynamic range imaging using differently-exposed image pair[J]. *Sensors*, 2017, 17(7): 1473.

- [49] Heo Y S, Lee K M, Lee S U. Depth maps and high-dynamic range image generation from alternating exposure multiview images[J]. *Journal of Electronic Imaging*, 2017, 26(4): 043006.
- [50] Pu Z Y, Guo P Y, Asif M S, et al. Robust high dynamic range (HDR) imaging with complex motion and parallax[M]//Ishikawa H, Liu C L, Pajdla T, et al. *Computer vision-ACCV 2020. Lecture notes in computer science*. Cham: Springer, 2021, 12623: 134-149.
- [51] Orozco R R, Loscos C, Martin I, et al. HDR multiview image sequence generation [M] // *High dynamic range video*. Amsterdam: Elsevier, 2017: 61-86.
- [52] Pizer S M, Johnston R E, Erickson J P, et al. Contrast-limited adaptive histogram equalization: speed and effectiveness[C]//*Proceedings of the First Conference on Visualization in Biomedical Computing*, May 22-25, 1990, Atlanta, GA, USA. New York: IEEE Press, 1990: 337-345.
- [53] Jobson D J, Rahman Z, Woodell G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes [J]. *IEEE Transactions on Image Processing*, 1997, 6(7): 965-976.
- [54] Liu J Y, Xu D J, Yang W H, et al. Benchmarking low-light image enhancement and beyond [J]. *International Journal of Computer Vision*, 2021, 129(4): 1153-1184.
- [55] Petschnigg G, Szeliski R, Agrawala M, et al. Digital photography with flash and no-flash image pairs[J]. *ACM Transactions on Graphics*, 2004, 23(3): 664-672.
- [56] Krishnan D, Fergus R. Dark flash photography[J]. *ACM Transactions on Graphics*, 2009, 28(3): 1-11.
- [57] Wang J, Xue T F, Barron J T, et al. Stereoscopic dark flash for low-light photography[C]//*2019 IEEE International Conference on Computational Photography (ICCP)*, May 15-17, 2019, Tokyo, Japan. New York: IEEE Press, 2019: 1-10.
- [58] Dong X, Li W X. Shoot high-quality color images using dual-lens system with monochrome and color cameras[J]. *Neurocomputing*, 2019, 352: 22-32.
- [59] Guo P Y, Ma Z. Low-light color imaging via dual camera acquisition[M]//Ishikawa H, Liu C L, Pajdla T, et al. *Computer vision-ACCV 2020. Lecture notes in computer science*. Cham: Springer, 2021, 12623: 150-167.
- [60] Liu Y. Special research report on optics: continuous innovation of mobile camera and upgrading of multi-directional support technology [R]. Xiamen: Great Wall Glory Securities Co., LTD, 2020.
刘亿. 光学专题研究报告: 手机摄像头持续创新, 多方位支撑技术升级[R]. 厦门: 长城国瑞证券, 2020.
- [61] Brady D J, Fang L, Ma Z. Deep learning for camera data acquisition, control, and image estimation[J]. *Advances in Optics and Photonics*, 2020, 12(4): 787-846.