

---

°硕士





中国科学院大学  
University of Chinese Academy of Sciences

# 硕士学位论文

## 基于深度学习的场景光照估计研究

作者姓名：程大川

指导教师：陈彦云 研究员

中国科学院软件研究所

学位类别：工学

学科专业：计算机图形学

培养单位：中国科学院软件研究所

2019 年 6 月



# **Deep Scene Illumination Estimation**

**A thesis submitted to the  
University of Chinese Academy of Sciences  
in partial fulfillment of the requirement  
for the degree of  
Master of Engineering  
in Computer Graphics**

**By**

**CHENG Dachuan**

**Supervisor: Professor Chen Yanyun**

**Institute of Software, Chinese Academy of Sciences**

**June, 2019**



## 中国科学院大学 学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日 期：

## 中国科学院大学 学位论文授权使用声明

本人完全了解并同意遵守中国科学院大学有关保存和使用学位论文的规定，即中国科学院大学有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：

日 期：

导师签名：

日 期：





## 摘 要

摘要

关键词： 关键词



## **Abstract**

**Keywords:** keywords



## 目 录

第1章 绪论 .....	1
1.1 选题的背景及意义 .....	1
1.2 国内外研究现状 .....	2
第2章 光照估计数据集 .....	9
2.1 引言 .....	9
2.2 全景图像简介 .....	9
2.3 构建HDR全景数据集 .....	13
2.4 构建光照估计数据集 .....	13
2.5 数据集对光照估计效果的影响分析 .....	13
2.6 讨论 .....	14
2.7 本章小结 .....	14
第3章 基于深度学习的光照估计方法 .....	15
3.1 引言 .....	15
3.2 相关工作 .....	15
3.3 问题求解范围 .....	15
3.4 光照分布的球形谐波表示 .....	15
3.5 卷积神经网络结构 .....	15
3.6 损失函数 .....	15
3.7 实验结果与评估 .....	15
3.8 深入研究光照估计网络 .....	15
3.9 讨论 .....	15
3.10 本章小结 .....	15
第4章 光照估计的应用 .....	17
4.1 引言 .....	17
4.2 电脑端应用 .....	17
4.3 移动端应用 .....	17
4.4 本章小结 .....	17
第5章 总结与展望 .....	19
5.1 本文工作总结 .....	19
5.2 未来工作展望 .....	19

参考文献 .....	21
作者简历及攻读学位期间发表的学术论文与研究成果 .....	27
致谢 .....	29

## 图形列表

1.1 光照估计的应用 .....	1
1.2 光照表示模型的示例 .....	3
1.3 使用特殊设备估计光照 .....	5
2.1 全景图的几种投影方式 .....	11





## 表格列表



## 符号列表

### 字符

Symbol	Description	Unit
$I_{ij}$	identity tensor	1

### 算子

Symbol	Description
$\nabla$	gradient operator

### 缩写

CFD	Computational Fluid Dynamics
-----	------------------------------



## 第1章 绪论

### 1.1 选题的背景及意义

光照估计（又称光照分布估计）是从已知的彩色图像信息中，预测、估计、恢复出整个场景的光照分布。该问题的输入通常是若干张彩色图片，或者是一段视频，有时已知的几何或材质信息也被用来辅助估计光照。场景的光照分布是指场景中各个方向的光照的颜色和强度。较为常见的光照分布表示方法包括高动态范围（High Dynamic Range, HDR）全景图、球形谐波（Spherical Harmonics, SH）函数表示、基于物理的Sun-Sky模型表示等。其中精度最高、使用比较广泛的是HDR全景图像，而在实时渲染领域使用较多的是光照的SH表示。基于物理的光照模型则多用于室外场景。

光照估计作为计算机图形学和计算机视觉的基础问题之一，有着广泛的实际应用场景。例如：基于图像的渲染（Image Based Rendering, IBR）、增强现实（Augmented Reality, AR）、电影后期制作、真实感虚实交互等。图1.1展示了光照估计的应用之一。光照估计也与这两个学科中的许多其它问题息息相关。例如：双向反射分布函数（BRDF）估计、场景几何重构、本征信息提取、图像增强，等等。高质量的光照估计结果通常能够为这些问题的解决带来很大的帮助。



图 1.1 光照估计的应用之一。使用单张图片估计场景的光照，并利用估计的光照渲染一个新的物体合成到图像中。可以看出使用估计光照渲染后的3D物体，与场景中的已有物体在视觉上较为一致。

**Figure 1.1 3D rendering under the predicted illumination. visual effect of the 3D rendering is in line with the original image.**

从有限的图像信息估计出整个场景的光照分布是一个复杂的问题。首先，图像的视野范围比较有限，例如一张视场角（FOV）为60°的照片所拍摄到的区域，在其对应的全景图中占比不足6%。此外，一幅图片是光照分布、场景几何结构、物体材质、摄相机参数等多个单位之间的复杂交互结果（公式 1.1）。

$$Image = ComplexInteraction(Light, Geometry, Material, Camera) \quad (1.1)$$

通过公式1.1可以看出，在其它三个信息未知的情况下，从图像（Image）反推出光照（Light）是一个严重的不适定（ill-posed）问题。不仅如此，在不同的条件下拍摄的彩色图像可能存在很多误差。例如图像中的过曝光/欠曝光区域、相机畸变、不正确的白平衡等。这些都会对光照估计造成一定程度的干扰，增加光照估计的难度。

研究者在光照估计问题上开展了一系列的研究。Debevec[1] 首次提出可以通过拍摄不同曝光下的镜面球体，来获取一张高动态范围的全景图片。 *Related limited methods*. 不难看出，这些方法或依赖特殊的探针、或依赖特殊的拍摄设备、或依赖额外的辅助信息与模型假设。均具有一定的局限性。

近年来，深度学习在多种计算机视觉问题上大放异彩，用于分割、检测、标识、分类的神经网络层出不穷。一些研究者尝试将深度学习应用在光照估计问题当中。 *Related learning based methods*. 不过，目前已有的深度学习方法也有一定的局限性。训练一个鲁棒的神经网络往往需要大量的数据，而目前用于光照估计问题的数据集比较有限，主要包括：大规模的低动态范围全景数据集（SUN360 *CITE*等）和中小规模特定场景的高动态范围全景数据集（Laval Indoor等 *CITE*）。这些数据集在规模和质量上很难同时到达训练深度神经网络的要求。

在这样的背景下，本文在光照估计的两个方向开展研究。其一是构建一个具有一定规模和质量光照估计的数据集。这样的数据集不仅能被用来训练更加鲁棒的光照估计网络，也可以被应用到其它多种相关的深度学习问题当中。其二是在已有数据集和本文构建的数据集基础上，深入探索基于深度学习的光照估计方法，对其中的网络结构，网络参数，损失函数，光照表示等多个模块进行细致的对比和研究。

## 1.2 国内外研究现状

### 1.2.1 光照的表示

光照的表示对光照估计问题至关重要。场景的光照分布有着多种多样的表示方法。其中高动态范围（High Dynamic Range，简称HDR）全景图是一种被广泛使用而且精度较高的光照表示方式。与普通的8位三通道图像不同，HDR图像的颜色值范围可以从0取到非常大。这意味着HDR图像可以更细致地表示每个像素位置的真实亮度值。因此无论是HDR全景图表示的光照，还是使用其渲染的物体，都能更接近真实值。Reinhard等人在 *High Dynamic Range Imaging*[2]中

对HDR图像进行了详细的介绍。

研究者们也提出了一些小巧、高效的光照近似模型，代价是牺牲一部分精度。Ramamoorthi和Hanrahan[3]提出使用球形谐波函数来表示场景的光照。这种方法表示的光照，不仅参数规模较小，而且其在渲染物体时非常高效。他在文章中指出，该模型仅使用9个系数时，在漫反射物体上渲染结果的平均误差也不超过1%。不过由于球形谐波函数本身的特性，这种方式不能很好的保留光照分布的高频细节。尽管如此，它的简洁和高效使得它成为实时渲染中的重要方式之一[4, 5]。许多光照估计方法中采用了这种模型来表示场景的光照分布。图1.2展示了不同的几种光照表示示例。



图 1.2 几种不同的光照表示模型对比

Figure 1.2 Comparison of different representations of illumination

在室外场景中，光线的主要来源是太阳光、光在空气中的散射折射、以及光在地面的反射。因此基于物理的光照表示模型更加适合表示室外场景。Perez等人[6]提出一个参数化的物理模型来表示天空中的发光度（Luminance）分布。之后的许多工作 [7–10] 考虑了空气散射和折射、大气浑浊度的影响，将该模型扩展为多种不同类型颜色模型。Hosek和Wilkie[11, 12]基于这些工作，考虑并加入了多种因素，整理出一个可调节太阳位置、大气浑浊度、地面反照率的物理参数模型。这类模型的参数量较小，精度也很可观，因此成为室外场景光照估计问题中常用的光照表示方法。

光照分布也有很多其它的表示方式。Ng等人[13]提出使用小波函数来近似光照分布。LeGendre等人[14]提出了一个实用的框架，能够使用LED灯台精确地重构出各个方向的场景光照。近期Weber[15]结合深度学习，使用自编码器（auto-encoder）对光照分布进行建模，为光照估计问题提供了新的思路。

### 1.2.2 使用光探测物估计光照

光照估计是一个复杂的问题，研究者们常常采用多种方法的组合来解决这个问题。使用光探测物（light probe）来估计光照就是这些方法之一。常见的光照探测物包括：镜面球体，漫反射球体，镜面反射/漫反射混合球体，人眼，人脸等等。Debevec[1]建立了一个基于光照模型的虚拟物体插入系统。其中所涉

及光照的就是使用镜面反射球体作为光探测物来估计的。该工作首次提出可以通过拍摄不同曝光下的镜面球体，来获取一张高动态范围的全景图片。在之后的工作中，Reinhard等人[2]和Debevec等人[16]指出，使用一个漫反射球体，或者漫反射与镜面反射表面混合的球体，也可以达到同样的效果。值得一提的是，Debevec在其工作[1]中，将远距离场景、近距离场景、以及待插入的虚拟物体分离开，并假设插入的虚拟物体并不会造成远距离场景的变化。之后的虚拟物体插入、增强现实等工作都遵从了这个假设。

上述的光探测物大多是已知物体几何和表面材质、且预先放置在场景中的规则物体。还有一些特殊的“物体”也可以作为光探测物。例如人的眼球，人脸，人手等常出现照片中的元素。Tsumra等人[17]在假设人眼是规则球体的前提下，利用眼球上光线的反射，估计场景的光照。Nishino和Nayar[18]分析了眼球的大致结构，并利用包含眼球的照片估计场景的光照分布。不过该方法需要分辨率和清晰度较高的相机，而且其文中也指出该方法没有考虑瞳孔颜色、虹膜颜色等因素。而这些因素都会对光照估计产生较大影响。人脸作为照片中经常出现的物体，也经常被用作光探测物。Wen等人[19]通过一张人脸照片，估计出光照的SH表示进而对人脸实现重照（relight）Wang等人[20]提出一种基于马尔科夫随机场的能量最小化框架，意图从正脸照片中恢复出人脸的形状、反射率和光照。在之后的工作中Shim[21]、Knorr等人[22]、Shahlaei和Blanz[23]探索了如何从人脸照片中估计较为精确的光照。Yao等人[24]使用普通相机和深度相机下的人体手部图像，通过人手的亮度和法线，估计出由若干个球形谐波系数近似的低频光照。

可以看出，这类估计光照的方法都需要一个已知其类型、几何、材质的光探测物存在于场景内。无论是规则的球体，还是人的各个部位，它们都需要照片中包含指定的物体和元素。而在大多数的光照估计问题中，很难保证这些标志物或探测物一定存在，这无疑限制了此类方法的应用。

### 1.2.3 使用特定设备估计光照

降低光照估计问题难度的另一个思路是使用特定的设备、装置、应用。Pilet等人[25]使用多个不同位置的相机和一个平面标定物构建了一个3D估计系统。通过追踪该平面标定物并计算其中的高光和阴影，估计出场景中的几何和光照。Yoo和Lee[26]提出了一个由鱼眼镜头、中性衰减片（Neural Density Filter, ND）和普通相机构成的光照探测系统。他们通过鱼眼镜头获取一个半球面，使用ND片直接感知明亮区域，并通过一系列算法实时地估计场景中的光



照。Cossairt[27]等人使用一组透镜、光场相机、光场映射器，构建了一个适合单光源较暗场景下的光照估计系统。随后Imai等人[28]使用多光谱成像设备，探究了可变亮度阈值、色调、偏振滤镜在检测光照条件较为复杂的场景下的高光分布。提出了适用于偏高光反射物体上的光照估计方法。Tocci等人[29]构建了一个光学设备，用以获取影视级别的高动态范围视频。Manakov等人[30]提供了一个插件类型的相机硬件，用以与高动态范围图像，多光谱、偏振和光场等相关的应用。Cailian等人[31]使用一个阴影探测器，来解决实时增强现实中的光照估计问题。Kán和Peter[32]通过全景图像拼接技术，建立了一个在智能相机中用于捕捉高动态范围全景图的应用。

不过，通过这些使用特殊设备或者特定应用获取场景光照的方法并不是严格意义上的光照估计。他们其实是倾向于解决快速获取高动态范围图像或视频的工程问题。图1.3展示了上述的一些的光照估计系统的使用情景，可以看出，



图 1.3 几种使用特殊设备估计场景光照的系统

Figure 1.3 Illumination estimation system using specific devices.

这种光照估计方法所需的设备通常比较复杂。虽然在实验室环境中能够有效运行，但是这些设备的类型和数量无疑提高了这种方法的应用门槛。

#### 1.2.4 使用额外信息估计光照

从单张图片恢复或估计出整个场景的光照分布是个复杂的问题。大部分的光照估计方法考虑借用额外的信息来辅助估计光照。例如，深度图像信息、场景几何形状、物体几何结构等。

深度信息对于光照估计问题有着很大的帮助。Knecht等人[33]借助深度相机和鱼镜头重建场景模型并提取了光源位置。Meilland等人[34]利用深度相机实时地创建稠密的高动态范围全景图，并使用K-means算法从环境贴图中提取点光源位置。Barron和Mailk[35]提出了一种从彩色图像和深度图像中提取本征信息的方法。他们使用具有噪声的深度图像来辅助估计场景的几何结构，并在这些信息的基础上进一步提取彩色图像中的本征信息。Zhang等人[36]使用RGBD相机，将拍摄到的室内场景建模为一个包含光源、材质和几何的没有杂物的房间模型。并提供了对该模型进行场景编辑的方法。

在已知场景或物体几何结构的情况下估计光照的分布也是主要的研究方向之一。例如，在形状类似桌子、平台、平板等结构的场景中，Li等人[37]的方法可以结合图片中的阴影、明暗、高光信息估计出多个方向的光源信息。Ramamoorthi和Hanrahan[38]使用物体的几何结构和多个视角下的照片作为输入，估计出照片中物体的表面材质并进一步计算所在场景的静态光照信息。不过该方法需要多种假设——曲线表面、各向同性BRDF、没有多重反射等。Sato等人[39]指出通过物体遮挡关系估计场景光照的有效性。他们通过分析给定几何的物体上的明亮区域和阴影区域，估计出较为真实的场景光照。在给定朗伯体（Lambertian）反射的物体几何时，Wang和Samaras[40]的方法可以从单张图片估计出多个方向的光照信息。Panagopoulos等人[41]提出了一个新的框架，来从单幅图片和粗糙的3D几何中，恢复出光照环境和估计场景中的投影阴影。该方法描述了一个高阶马尔可夫随机场（MRF）照明模型。该模型将低阶阴影证据与高阶先验知识相结合，用于投影阴影和照明环境的联合估计。

此外还有一些方法在已知物体类型的情况下，利用图片中物体的先验或者假设知识来帮助估计光照。例如前文提到的Nishino和Nayar[18]假设人的眼球是规则的球体，并基于此假设从人眼估计场景的光照。类似的方法[42–44]虽然不需要预先知道精确的场景几何结构，但都要通过图像中物体的类型得到一些先验知识假设和约束。进而分析、预测、估计、恢复出场景光照。

增加彩色图片的数量也会对光照估计有很大帮助。Sato等人[45]使用两张全方位的照片构建出场景的大致几何，然后根据不同快门速度拍摄的全方向图像序列计算场景的辐射度，并将其映射到构建的几何模型上。这个映射到几何上的辐照度分布就可以作为光照来渲染新的3D物体。Nishino等人[46, 47]在已知物体的大致几何的情况下，分别对发光体和普通物体拍摄多张图片来估计场景光照。Yu等人[48]则通过多视角图片来恢复出固定光照条件下的物体纹理和光照分布。Wu等人[49]建立了一个纯基于图片的形状、表面、光照的估计模型。Shan等人[50]在解决重构场景的问题时，提出了一种从大规模不同光照条件图片集中估计结点反照率和光照参数的方法。受到该方法的启发，Lalonde和Matthews[51]在同一个室外建筑物的不同图片中恢复出每幅图片对应的光照条件。

需要注意的是，以上几种增加辅助信息的方式并不是互斥的，研究者们可以选择使用多种额外信息共同辅助估计光照。例如Marschner[52]在已知物体几何的前提下，使用一组照片估计出物体表面的反射情况，进而估计出场景光照

信息。Haber等人[53]利用一个物体在多个角度下的图片和已知几何结构估计其光照分布。使用一种或多种额外的信息来辅助光照估计一直是解决光照估计问题的主要方法，也是能够提升光照估计的效果的主要方式。但是此类方法通常需要使用额外的输入设备（例如深度相机），繁琐的获取步骤（例如多次拍摄），以及一些先验知识。而且使用的额外信息越多，所需的设备就越多，步骤就越繁琐，这难免限制了它们的应用场景。

### 1.2.5 使用简化的光照模型估计光照

光照估计是一个已知条件较少，求解结果复杂，涉及因素繁多的问题。除了使用额外的设备增加输入信息规模的方法外，降低该问题难度的另一个思路是简化光照的表示。从有限的信息中估计出复杂的场景光照分布是比较困难的，所以许多方法尝试选取相对简单的光照近似模型表示光照，即通过牺牲一部分精度来降低待估计参数量的规模，从而达到简化光照估计问题的目的。其中最常见的使用球形谐波（spherical harmonic）函数近似的光照模型。该模型最早由Ramamoorthi[3]在2001年提出。通过应用该模型，光照的低频部分可以使用少量的系数（通常为9-16组，约27-48个）来近似。虽然这种表示方式对光照分布中的高频细节不太友好，但在渲染常见的漫反射物体时却有着极小的误差。因此许多工作[22, 38, 42, 44, 54–56]通过估计少量的SH系数来达到估计光照的目的。与之类似，Barron和Malik[57]在光照估计问题中使用小波函数来近似场景光照，并将其与SH表示进行了对比，指出了在表示光照时球形谐波函数相较于小波函数的优点。早期的一些光照估计工作[37, 39–41, 45]将光照分布简化为若干个点光源的集合。进而将光照分布估计问题转化为预测光源数量、位置和大小的问题。预测这种类型的光照比较简单，但遗憾的是真实场景中的光照分布往往比较复杂，能够使用这种类型表示的场景并不是很多。

对于特定的场景（比如室外），基于物理模型的光照表示是一个很好的选择。这种模型往往能够使用很小的参数量近似出精度较高的光照分布。其中最具有代表性的是Perez[6]在1993年提出的天空的发光度分布模型。该模型经过多次补充，修改和完善[7–12]，目前已经成为了一个包含太阳位置、大气浑浊度、地面反照率等多个具体物理意义参数的物理模型。因此大部分室外光照估计方法[58–61]都采用了这类模型。

值得注意的是，几乎所有的光照估计算法，都对要估计的光照进行一定的简化。选取合理的光照分布简化方法对于光照估计问题至关重要，需要综合考量已知条件，应用需求，核心方法框架等。

#### 1.2.6 基于图像分析估计光照

[62] [63]

#### 1.2.7 用户交互辅助估计光照

#### 1.2.8 深度学习与光照估计的结合

#### 1.2.9 光照估计相关数据集

深度学习与光照估计的结合是主流趋势之一。

## 第2章 光照估计数据集

### 2.1 引言

对于深度学习任务来说，训练数据的规模、质量对网络最终的表现有着直接的影响。小规模的数据、低质量的数据、不平衡的数据都可能会导致网络训练失败。从图片估计光照是一个非常复杂的问题，小规模低质量的数据往往很难训练出较好的预测网络。目前用于光照估计问题的数据集比较有限，主要包括大规模的低动态范围（LDR）全景图和小规模的高动态范围（HDR）全景图。这两类数据集都在规模和质量上，无法同时满足训练一个鲁棒的光照估计网络的条件。因此本文将通过收集、拍摄、筛选、整理等多个严格细致的步骤，构建一个同时包含多类场景的、多种光照条件，而且具有一定规模的光照估计数据集。

本章主要介绍该数据集的构建方式以及在该数据集上的对比试验。本章内容中，首先对低动态范围全景图像和高动态范围全景图像做出介绍；然后阐述HDR全景图的主要拍摄与合成方法；接着展示了用HDR全景图像构建光照估计数据集的若干方法，最后在该数据集上进行多项对比实验，给出了数据集质量和规模对于光照估计问题的影响。

### 2.2 全景图像简介

全景图（panorama）是一种广角图，可以以画作、照片、影片、三维模型的形式存在。全景图这个词最早由爱尔兰画家罗伯特·巴克提出，用以描述他创作的爱丁堡全景画。现代的全景图多指通过相机拍摄并在计算机上加工而成的图片[64]。全景图存储了以相机位置为中心的每个角度的颜色信息，颜色信息与普通图片类似，常用RGB三个通道分别存储。全景图根据其中的颜色数值范围，可分为低动态范围全景图和高动态范围全景图。

#### 2.2.1 获取

拍摄全景图像的方式主要有两类。一类是使用专业的全景相机拍摄设备，这类设备大多由数个鱼镜头形式的广角相机组成。在拍摄时，设备中的相机使用相同的相机参数同时拍摄，随后内置的固件或软件会对所拍摄到的图像进行投影变换、校正、拼接，形成一张全景图。另外一类是使用普通相机和相机

旋转装置，对多个角度拍摄，随后在手动利用拼接算法或拼接软件将这些图像连接到一起。由于这种方式拍摄到的图片并不在同一个时刻，所以需要保证场景中不能包含过多的快速运动的物体。此外，大多数的现代智能手机都提供了手动拍摄“全景图”的方式。不过需要注意的是，由于相机视角范围限制，以人体为轴旋转手机相机拍摄到的“全景图”多称为“宽景图”，因为顶部视角和底部视角的区域仍然有很大缺失。

### 2.2.2 投影方式

全景图像的存储需要考虑投影方式和颜色的数值类型。在全景图像中，以相机为中心的视场可以被视为一个球体的表面，因此在存储和浏览全景图时，需要将全景球表面投影在二维表面中。常见的投影方式有等距投影，圆柱投影，球形投影，立方体贴图投影，立体投影等。

- **圆柱投影** (cylindrical projection) 该投影方式是将全景球置于其外切圆柱内，并由球心向圆柱面做投影，随后将圆柱内表面横向展开后的图像即为球形全景图的圆柱投影。这种投影下的全景图在两极会发生无限的纵向拉伸，因此圆柱投影后的图像无法包含靠近两极的信息，也即这种投影方式无法表示垂直视角为 $180^\circ$ 的全景图。柱面投影是传统摆动镜头全景胶片相机所提供的标准投影方式。相对于全角度的全景图，该投影方式更适合在垂直视角小于 $120^\circ$ 的宽景图，常用于现代智能手机的全景图预览。

- **等距投影** (equirectangular projection), 也称等距圆柱投影。该投影方式是将球面的经度和纬度坐标线性变换为图像空间的横纵坐标。经过投影处理后的全景图像是一幅宽高比例为2:1的图片。这种投影的特点是越接近两级，图像的变形就越严重。投影后的全景图在预览时一目了然，而且这种投影方式较为简单，是存储和预览全景图最常见的方式之一。

- **圆周鱼眼投影** (circumferential fisheye projection), 也称圆形投影或镜面球投影，为角投影的一种。这样的投影图像看起来像一个用超级圆形鱼镜头所拍摄的图片，虽然这种投影方式依然覆盖 $360^\circ$ 的视角，但是其边缘却被极端扭曲和变形。这种投影方式常见于全景图拍摄设备中。全景相机一般由两个朝向相反的 $180^\circ$ 视角鱼镜头。每个镜头所拍摄到的图片均为一个垂直和水平视角各 $180^\circ$ 的圆周鱼眼投影视图。

- **立方体贴图投影** (cubemap projection) 在计算机图形学中，立方体投影是常用的环境映射方法之一，常用于游戏场景中的天空盒，相当于等距柱状投影的优化版，环境投影到立方体之后可分六个正方形纹理来存储；或者将立方



体展开，存储于一个纹理中的六个区域内。在全景图像及视频中，立方体投影将球形视频映射到它的外接立方体上，立方体的上下两个面分别对应两极区域，中间的四个面对应赤道区域。

- **立体投影**（stereographic projection），即常见的小行星投影方式。



图 2.1 全景图几种投影方式

Figure 2.1 projection of panorama

除此之外，还有墨卡托投影、正弦投影等、直线投影用于不同的领域。图2.1展示了几种常见的投影方式。


### 2.2.3 动态范围


图像的动态范围（dynamic Range）是指一个图像中最亮和最暗部分之间的相对比值。[64]。根据动态范围的大小可以将图像分为低动态范围（或称为标准动态范围）图像和高动态范围图像。传统的8位图像将颜色值存储为[0, 255]范围内的整数，低动态范围图像中的颜色只能在这256个数中取值。图图片展示了同一场景不同曝光条件下的照片，可以看出每幅图片都有一定的过曝和欠曝区域。例如在过曝图片中，太阳和其周围区域的颜色均为白色，但实际上太阳的亮度要远远高过天空的亮度；同样的，在欠曝区域中，虽然大部分区域同为黑色，但实际上这些区域的明暗程度也是千差万别的。


而相比低动态范围的图像，高动态范围图像可以提供更多的动态范围（理论上是0到无限大）。高动态范围全景图也是具有这种动态范围的全景图像。高动态范围全景图的动态范围可以高达 $2^{32}$ ，而人类的眼睛所能看到的范围是 $10^5$ 左右[64]。因此高动态范围全景图像可以作为真实的光照信息，这是低动态范围全景图不能比拟的。图片展示了两组使用低动态范围全景图和高动态范围全景图的渲染结果，可以看出，使用低动态范围全景图渲染的结果颜色值对比平缓，看起来很不真实。而使用高动态范围全景图作为光照时，渲染结果包含了足够的对比度和锐利的阴影。图片中的第二组使用了包含更亮光源的场景作为对比，它们之间的差异更加明显。下文中的LDR将代指低动态范围全景图，HDR将代指高动态范围全景图。

## 2.2.4 HDR的获取与存储

HDR全景图的获取方式与普通全景图类似，不过由于HDR包含了更大的动态范围。而普通相机单词拍摄无法满足这个要求，因此需要使用对每个视角的图像都要使用不同的曝光值来进行拍摄。HDR全景图的获取分为多曝光拍摄，多视角拍摄，全景图拼接，清洗和筛选，曝光融合五个步骤。

- **多曝光拍摄。**HDR图像通常无法直接由拍摄设备直接获取，因此需要利用不同的曝光值对同一场景多次拍摄。曝光值通常由相机的ISO、快门速度、光圈大小共同决定。图展示了在不同曝光条件下拍摄的图像。

- **多视角拍摄。**单个相机的视角有限，在合成全景图之前，需要对场景的不同视角拍摄。为了保证拼接后的图片质量，拍摄时相机的镜头不能有太大的平移。因此拍摄全景图时，相机位置多由精密的机械装置自动控制。图展示了一个用于拍摄全景图的机械装置。此外需要注意的是，对于每个拍摄视角，拍摄时的曝光值序列和相机的其它参数（例如白平衡，相机视场范围等）要完全一致，以保证在全景图拼接时的各个视角图片的一致性。

- **全景图拼接。**使用普通相机对多个视角进行拍摄后，需要将这些图片拼接到一起。这通常需要做图片间的特征匹配等等，目前全景图的拼接有较为成熟的算法和工具，这里不再展开叙述。另外，目前市面上出现了一种专门用于拍摄全景图的全景相机，其中有些全景相机支持通过调节ISO，快门速度等拍摄多种曝光条件下的全景图，例如小米全景相机[65]等。使用这种相机可以省去图像拼接的步骤，每次拍摄只需要关注相机的曝光即可，也可以避免拼接时产生的瑕疵结果。图展示了使用全景相机拍摄全景图的方式之一。

- **清洗和筛选。**拍摄结果中不可避免的会有一些质量较差的图片，因此需要进行筛选和清理。质量较差的图片通常包括：图片中场景不一致（经常由移动物体在不同时刻被拍摄导致）、图片色差过大、图片噪点过多（经常由较暗场景中用过大的ISO导致）、相机本身的拍摄异常等。这些都需要人工观察、筛选、清理和重拍，以保证最终合成的HDR全景图的质量。

- **曝光HDR融合。**当获得多张不同曝光的全景图后需要将他们融合为一张HDR图，常见的融合算法有基于基于信息熵的算法、基于双边滤波的算法、基于亮度梯度大小的算法、基于拉普拉斯金字塔的算法等。曝光融合工具中，PTGUI[66]能够提供很好的融合结果。

曝光融合后的全景图就是一幅HDR全景图，这种全景图包含了很高的动态范围，可以作为场景真实光照的表示。HDR全景图像的投影方式和普通全景图完全



一致，它们之间的区别只是每个像素位置的数值类型和数值范围不同，常见的HDR图像保存格式有TIFF, HDR, RGBE, EXR等。

## 2.3 构建HDR全景数据集

本节主要介绍构建HDR全景数据集的方法与步骤。构建时需要考虑数据的多样性，保证数据质量。

### 2.3.1 场景选择

HDR全景数据集需要考虑到场景的多样性，为此，本文选取了多个场景进行拍摄，包含了室内、室外，森林，公园，公寓，小区，建筑群等多种常见场景，晴天，阴天，多云等多种气象条件，清晨、中午、下午、傍晚、夜间等多种拍摄时间，以及春夏秋冬四季多个拍摄季节，其中各场景比例如表格表格所示。

### 2.3.2 拍摄设备选择

拍摄HDR全景图的方式有两种，一种是使用普通相机多次拍摄并进行拼接，另一种是直接使用全景相机拍摄，本文采用的的方式是后者，即直接使用全景相机拍摄，这样可以避免大量的拼接操作，只需要根据场景调整不同的曝光范围即可。

### 2.3.3 曝光融合工具

在进行曝光融合时，通过多种曝光融合工具的对比，发现PTGUI[66]能够提供很好的融合结果，因此本文在构建该光照估计数据集时主要使用此工具进行曝光融合，融合时的参数如表格表格所示。

通过如上步骤，我们共拍摄了约5000张全景图像，经过清理和筛选，共合成了约550组HDR全景图。图图片是展示了部分HDR全景图，可以看出该数据集包含了多种场景和光照条件，而且图片质量较高，几乎没有早点和拼接痕迹。

## 2.4 构建光照估计数据集

### 2.5 数据集对光照估计效果的影响分析

为了验证数据集在光照估计中的有效性，本文设计另一个基础的光照估计深度学习网络。并在该网络上使用多种数据集进行训练和测试，

### 2.5.1 用于对比的光照估计网络

从图片预测整个HDR图往往比较困难，为了有效地在该数据集上进行试验，本文搭建了一个从图片到球形谐波系数的网络。

### 2.5.2 数据集划分

对比试验主要分为三个部分，首先是对比不同的数据规模对光照估计网络的影响，其次是对比数据的多样性大小对光照估计网络的影响，最后是本文构建的数据集和部分已有数据集的对比。在前两部分实验中，所有的数据均为本文所构建的数据集，因此将该数据集按照训练集400、测试集100、验证集50的比例划分后，在训练集上训练，以验证集上的结果作为实验结果进行分析对比和验证。在与其它数据集对比的实验中，为了公平实验，用作对比和分析的数据均由真实的图片和对应的HDR全景图作为对比，而本文所构建的数据均用于训练光照估计网络。此外，在数据划分时，是在HDR全景数据上进行划分后，再在每一幅HDR全景图中提取视角图片，而不是在提取的图片中进行划分，这样可以保证同一幅HDR全景图不会同时出现再不同的数据集中。

### 2.5.3 数据的规模对光照估计的影响

### 2.5.4 数据多样性对光照估计的影响

### 2.5.5 与其它数据集的对比

本文与其它数据集在两个方面进行对比，一个是

## 2.6 讨论

可以看出，HDR全景图的拍摄需要对同一场景多次拍摄，因此快速变动的场景。不过对于光照估计问题来说，该数据集的多样性已经足够规模依然可以扩充，地点可以增多，动态范围可以继续增加。

## 2.7 本章小结

## 第3章 基于深度学习的光照估计方法

- 3.1 引言
- 3.2 相关工作
- 3.3 问题求解范围
- 3.4 光照分布的球形谐波表示
- 3.5 卷积神经网络结构
- 3.6 损失函数
- 3.7 实验结果与评估
- 3.8 深入研究光照估计网络
- 3.9 讨论
- 3.10 本章小结



## 第4章 光照估计的应用

- 4.1 引言
- 4.2 电脑端应用
- 4.3 移动端应用
- 4.4 本章小结



## 第5章 总结与展望

### 5.1 本文工作总结

### 5.2 未来工作展望





## 参考文献

- [1] DEBEVEC P. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography[C]//Proceedings of the 25th annual conference on Computer graphics and interactive techniques. ACM, 1998: 189-198.
- [2] REINHARD E, WARD G, PATTANAIK S, et al. High dynamic range imaging: Acquisition, display, and image-based lighting[M]. Elsevier, 2005.
- [3] RAMAMOORTHY R, HANRAHAN P. An efficient representation for irradiance environment maps[C]//Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, 2001: 497-500.
- [4] GREEN R. Spherical harmonic lighting: The gritty details[C]//Archives of the Game Developers Conference: volume 56. 2003: 4.
- [5] SLOAN P P. Stupid spherical harmonics (sh) tricks[C]//Game developers conference: volume 9. 2008: 42.
- [6] PEREZ R, SEALS R, MICHALSKY J. All-weather model for sky luminance distribution—preliminary configuration and validation[J]. Solar energy, 1993, 50(3):235-245.
- [7] NISHITA T, DOBASHI Y, NAKAMAE E. Display of clouds taking into account multiple anisotropic scattering and sky light[C]//Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM, 1996: 379-386.
- [8] SIRAI T N T, NAKAMAE K T E. Display of the earth taking into account atmospheric scattering[C]//Siggraph: volume 93. Citeseer, 1993: 175.
- [9] PREETHAM S. A practical analytic model for daylight[J]. 1999.
- [10] RAAB M, SEIBERT D, KELLER A. Unbiased global illumination with participating media [M]//Monte Carlo and Quasi-Monte Carlo Methods 2006. Springer, 2008: 591-605.
- [11] HOSEK L, WILKIE A. An analytic model for full spectral sky-dome radiance[J]. ACM Transactions on Graphics (TOG), 2012, 31(4):95.
- [12] HOŠEKHOŠEK L, WILKIE A. Adding a solar-radiance function to the hošek-wilkie skylight model[J]. IEEE computer graphics and applications, 2013, 33(3):44-52.
- [13] NG R, RAMAMOORTHY R, HANRAHAN P. All-frequency shadows using non-linear wavelet lighting approximation[C]//ACM Transactions on Graphics (TOG): volume 22. ACM, 2003: 376-381.
- [14] LEGENDRE C, YU X, LIU D, et al. Practical multispectral lighting reproduction[J]. ACM Transactions on Graphics (TOG), 2016, 35(4):32.
- [15] WEBER H, PRÉVOST D, LALONDE J F. Learning to estimate indoor lighting from 3d objects[C]//2018 International Conference on 3D Vision (3DV). IEEE, 2018: 199-207.

- [16] DEBEVEC P, GRAHAM P, BUSCH J, et al. A single-shot light probe[C]//ACM SIGGRAPH 2012 Talks. ACM, 2012: 10.
- [17] TSUMURA N, DANG M N, MAKINO T, et al. Estimating the directions to light sources using images of eye for reconstructing 3d human face[C]//Color and Imaging Conference: volume 2003. Society for Imaging Science and Technology, 2003: 77-81.
- [18] NISHINO K, NAYAR S K. Eyes for relighting[J]. ACM Transactions on Graphics (TOG), 2004, 23(3):704-711.
- [19] WEN Z, LIU Z, HUANG T S. Face relighting with radiance environment maps[C]//2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.: volume 2. IEEE, 2003: II-158.
- [20] WANG Y, LIU Z, HUA G, et al. Face re-lighting from a single image under harsh lighting conditions[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.
- [21] SHIM H. Faces as light probes for relighting[J]. Optical Engineering, 2012, 51(7):077002.
- [22] KNORR S B, KURZ D. Real-time illumination estimation from faces for coherent rendering [C]//2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2014: 113-122.
- [23] SHAHLAEI D, BLANZ V. Realistic inverse lighting from a single 2d image of a face, taken under unknown and complex lighting[C]//2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG): volume 1. IEEE, 2015: 1-8.
- [24] YAO Y, KAWAMURA H, KOJIMA A. The hand as a shading probe[C]//ACM SIGGRAPH 2013 Posters. ACM, 2013: 108.
- [25] PILET J, GEIGER A, LAGGER P, et al. An all-in-one solution to geometric and photometric calibration[C]//2006 IEEE/ACM International Symposium on Mixed and Augmented Reality. IEEE, 2006: 69-78.
- [26] YOO J D, LEE K H. Real time light source estimation using a fish-eye lens with nd filters[C]// 2008 International Symposium on Ubiquitous Virtual Reality. IEEE, 2008: 41-42.
- [27] COSSAIRT O, NAYAR S, RAMAMOORTHI R. Light field transfer: global illumination between real and synthetic objects[C]//ACM Transactions on Graphics (TOG): volume 27. ACM, 2008: 57.
- [28] IMAI Y, KATO Y, KADOI H, et al. Estimation of multiple illuminants based on specular highlight detection[C]//International Workshop on Computational Color Imaging. Springer, 2011: 85-98.
- [29] TOCCI M D, KISER C, TOCCI N, et al. A versatile hdr video production system[C]//ACM Transactions on Graphics (TOG): volume 30. ACM, 2011: 41.
- [30] MANAKOV A, RESTREPO J, KLEHM O, et al. A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging[J]. ACM Transactions on Graphics, 2013, 32(4):47-1.

- 
- [31] CALIAN D A, MITCHELL K, NOWROUZEZAHRAI D, et al. The shading probe: Fast appearance acquisition for mobile ar[C]//SIGGRAPH Asia 2013 Technical Briefs. ACM, 2013: 20.
- [32] KÁN P. Interactive hdr environment map capturing on mobile devices.[C]//Eurographics (Short Papers). 2015: 29-32.
- [33] KNECHT M, TRAXLER C, MATTAUSCH O, et al. Reciprocal shading for mixed reality[J]. Computers & Graphics, 2012, 36(7):846-856.
- [34] MEILLAND M, BARAT C, COMPORT A. 3d high dynamic range dense visual slam and its application to real-time object re-lighting[C]//2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2013: 143-152.
- [35] BARRON J T, MALIK J. Intrinsic scene properties from a single rgb-d image[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 17-24.
- [36] ZHANG E, COHEN M F, CURLESS B. Emptying, refurnishing, and relighting indoor spaces [J]. ACM Transactions on Graphics (TOG), 2016, 35(6):174.
- [37] LI Y, LU H, SHUM H Y, et al. Multiple-cue illumination estimation in textured scenes[C]//Proceedings Ninth IEEE International Conference on Computer Vision. IEEE, 2003: 1366-1373.
- [38] RAMAMOORTHY R, HANRAHAN P. A signal-processing framework for inverse rendering [C]//Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, 2001: 117-128.
- [39] SATO I, SATO Y, IKEUCHI K. Illumination from shadows[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(3):290-300.
- [40] WANG Y, SAMARAS D. Estimation of multiple illuminants from a single image of arbitrary known geometry[C]//European conference on computer vision. Springer, 2002: 272-288.
- [41] Panagopoulos A, Wang C, Samaras D, et al. Illumination estimation and cast shadow detection through a higher-order graphical model[C/OL]//CVPR 2011. 2011: 673-680. DOI: [10.1109/CVPR.2011.5995585](https://doi.org/10.1109/CVPR.2011.5995585).
- [42] BARRON J T, MALIK J. Shape, illumination, and reflectance from shading[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(8):1670-1687.
- [43] LOPEZ-MORENO J, HADAP S, REINHARD E, et al. Compositing images through light source detection[J]. Computers & Graphics, 2010, 34(6):698-707.
- [44] REMATAS K, RITSCHER T, FRITZ M, et al. Deep reflectance maps[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4508-4516.
- [45] SATO I, SATO Y, IKEUCHI K. Acquiring a radiance distribution to superimpose virtual objects onto a real scene[J]. IEEE transactions on visualization and computer graphics, 1999, 5(1):1-12.
- [46] NISHINO K, ZHANG Z, IKEUCHI K. Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis[C]//Proceedings

- Eighth IEEE international conference on computer vision. ICCV 2001: volume 1. IEEE, 2001: 599-606.
- [47] NISHINO K, IKEUCHI K, ZHANG Z. Re-rendering from a sparse set of images[J]. Department of Computer Science, Drexel University, 2005.
  - [48] YU T, WANG H, AHUJA N, et al. Sparse lumigraph relighting by illumination and reflectance estimation from multi-view images[C]//ACM SIGGRAPH 2006 Sketches. ACM, 2006: 175.
  - [49] WU C, WILBURN B, MATSUSHITA Y, et al. High-quality shape from multi-view stereo and shading under general illumination[C]//CVPR 2011. IEEE, 2011: 969-976.
  - [50] SHAN Q, ADAMS R, CURLESS B, et al. The visual turing test for scene reconstruction[C]//2013 International Conference on 3D Vision-3DV 2013. IEEE, 2013: 25-32.
  - [51] LALONDE J F, MATTHEWS I. Lighting estimation in outdoor image collections[C]//2014 2nd International Conference on 3D Vision: volume 1. IEEE, 2014: 131-138.
  - [52] MARSCHNER S R, GREENBERG D P. Inverse lighting for photography[C]//Color and Imaging Conference: volume 1997. Society for Imaging Science and Technology, 1997: 262-265.
  - [53] HABER T, FUCHS C, BEKAER P, et al. Relighting objects from image collections[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 627-634.
  - [54] KEMELMACHER-SHLIZERMAN I, BASRI R. 3d face reconstruction from a single image using a single reference face shape[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 33(2):394-405.
  - [55] GARRIDO P, VALGAERTS L, WU C, et al. Reconstructing detailed dynamic face geometry from monocular video.[J]. ACM Trans. Graph., 2013, 32(6):158-1.
  - [56] LI C, ZHOU K, LIN S. Intrinsic face image decomposition with human face priors[C]//European Conference on Computer Vision. Springer, 2014: 218-233.
  - [57] OKABE T, SATO I, SATO Y. Spherical harmonics vs. haar wavelets: Basis for recovering illumination from cast shadows[C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.: volume 1. IEEE, 2004: I-I.
  - [58] LALONDE J F, NARASIMHAN S G, EFROS A A. What does the sky tell us about the camera?[C]//European conference on computer vision. Springer, 2008: 354-367.
  - [59] LALONDE J F, NARASIMHAN S G, EFROS A A. What do the sun and the sky tell us about the camera?[J]. International Journal of Computer Vision, 2010, 88(1):24-51.
  - [60] LALONDE J F, EFROS A A, NARASIMHAN S G. Estimating the natural illumination conditions from a single outdoor image[J]. International Journal of Computer Vision, 2012, 98(2):123-145.
  - [61] SUNKAVALLI K, ROMEIRO F, MATUSIK W, et al. What do color changes reveal about an outdoor scene?[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1-8.

- [62] KARSCH K, SUNKAVALLI K, HADAP S, et al. Automatic scene inference for 3d object compositing[J]. ACM Transactions on Graphics (TOG), 2014, 33(3):32.
- [63] LOMBARDI S, NISHINO K. Reflectance and illumination recovery in the wild[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(1):129-141.
- [64] WIKIPEDIA. panorama[EB/OL]. 2019. <https://www.wikipedia.org/>.
- [65] XIAOMI. Xiaomi Panoramic Camera[EB/OL]. 2016. <https://www.mi.com/mj-panorama-camera/>.
- [66] PTGUI. new hourse internet services b.v, holland[EB/OL]. 2000. <https://www.ptgui.com/>.



## 作者简历及攻读学位期间发表的学术论文与研究成果

### 作者简历

#### 作者基本情况

程大川，男，汉族人，1994年11月11日生，山东菏泽人

2012年2月–2016年6月 山东科技大学信息学院软件工程工学学士

2016年9月–2019年6月 中国科学院大学软件研究所计算机应用技术工学硕士

#### 联系方式

通讯地址：北京市海淀区中关村南四街4号中国科学院软件研究所

计算机科学国家重点实验室

邮政编码：100190

电子邮箱：chengdc@ios.ac.cn

#### 已发表(或正式接受)的学术论文:

[1] Cheng, Dachuan, Jian Shi, Yanyun Chen, Xiaoming Deng, and Xiaopeng Zhang. "Learning Scene Illumination by Pairwise Photos from Rear and Front Mobile Cameras." In Computer Graphics Forum, vol. 37, no. 7, pp. 213-221. 2018.





## 致 谢

致谢

