

基于深度学习的逆向反射模型

王晰, 简振雄, 任明俊*

上海交通大学机械与动力工程学院机械系统与振动国家重点实验室, 上海 200240

摘要 为提高光度立体视觉技术处理各向同性非朗伯反射的能力, 提出一种基于深度学习的逆向反射模型, 通过提取与方位角差相关的图像特征, 弥补共位光源逆向反射模型的理论不足, 实现表面法向量高精度估计。该模型由三阶段子网络组成, 分别是方位角差子网络、逆向反射模型子网络与法向量估计子网络, 其中: 第一阶段子网络与第二阶段子网络共同实现像素值到法向量以及入射光线方向点积的高精度映射; 第三阶段子网络充分利用前两个子网络提取的特征, 实现表面法向量高精度估计。仿真实验表明, 所提方法对 100 种典型各向同性非朗伯反射均具有较好的处理能力; 基于标准数据集的真实实验证明, 所提方法能够取得平均 5.90° 的法向量估计精度, 充分证明所提方法的有效性。

关键词 光计算; 深度学习; 非朗伯反射; 光度立体视觉

中图分类号 TP399 **文献标志码** A

DOI: 10.3788/AOS230615

1 引言

光度立体视觉技术是计算机视觉领域的经典技术, 该技术利用场景在不同位置光源照射下的反射信息来估计物体表面法向量, 可用于磨料表面精细重建^[1]、机加工表面纹理测量、缺陷检测、微织物形态重建^[2]、点云拼接与大尺寸工件原位测量^[3]等。早期光度立体视觉技术^[4-6]假设物体表面为朗伯漫反射, 采用最小二乘法实现表面法向量估计, 然而真实物体表面通常展现复杂的非朗伯反射, 这使得早期光度立体视觉技术失效。为打破朗伯漫反射假设的束缚, 相关领域开展非朗伯光度立体视觉研究^[7], 旨在平行光源照射以及正交相机拍摄条件下, 消除各向同性非朗伯反射对法向量估计精度的影响。

非朗伯光度立体视觉方法包括传统参数法与数据驱动法。传统参数法采用鲁棒性估计法^[8-10]、参数化反射模型法^[11-14]与通用反射性质法^[15-16]构造参数化目标方程, 采用数值优化方法实现非朗伯反射表面法向量的高精度估计。该类方法本质上利用不同数学手段, 对非朗伯反射行为进行参数化建模, 消除非朗伯反射对法向量估计精度的影响。然而, 上述方法的泛化性有限, 仅能有效处理近似漫反射、小高亮反射等简单各向同性非朗伯反射。

数据驱动法旨在直接建立观测像素值与表面法向量之间的映射关系, 即逆向反射模型, 包括基于例子的

光度立体视觉技术^[17-18]与基于深度学习的光度立体视觉技术^[19-20]。Hertzmann 等^[17]提出方向一致性线索, 即两个具有相同表面法向量的点向相同的观察方向反射相同的光源能量, 并利用参考球面, 以搜索方式实现表面法向量估计。Hui 等^[18]利用 Mitsubishi Electric Research Laboratories (MERL)^[21]真实反射数据集构造虚拟参考球面来代替真实参考球面, 使该类技术有效应用于现实场景。Enomoto 等^[22]提出预计算策略, 极大提高了基于例子的光度立体视觉算法的计算速度。尽管相比于传统参数法, 基于例子的光度立体视觉算法有效提高了表面法向量的估计精度, 但是该类方法需要离散化法向量求解空间, 难以获得法向量最优解。

基于深度学习的光度立体视觉技术是当前光度立体视觉领域的热点方向, 该类方法旨在利用深度学习技术建立逆向反射模型。Santo 等^[23-24]首先利用当代深度学习技术, 根据方向一致性线索, 构建像素值向量与表面法向量之间的逆向映射关系, 并证明所提深度学习技术能够有效提高法向量的估计精度。但是, 该网络需要固定入射光线方向, 为了突破该限制, 相关研究提出两种经典的 convolutional neural network (CNN) 结构, 分别是 CNN based photometric stereo (CNN-PS)^[20]与 fully convolutional network for photometric stereo (PS-FCN)^[19]。

CNN-PS^[20]同样遵循方向一致性线索, 并设计了

收稿日期: 2023-03-02; 修回日期: 2023-05-26; 录用日期: 2023-06-13; 网络首发日期: 2023-06-28

基金项目: 国家重点研发计划变革性专项课题(2019YFA0706701)、国家自然科学基金(52175477)、国家自然科学基金合作研究项目(U22A20176)

通信作者: *renmj@sju.edu.cn

观测值图谱,按照特定顺序排列单个像素在不同入射光线方向下的像素值,随后利用CNN结构实现表面法向量估计。当光源数量较少时,观测值图谱记录的有效反射信息减少,导致法向量估计精度急剧下降。为了解决这一问题,Li等^[25]提出连接表适应稀疏光源,Zheng等^[26]利用插值网络将稀疏输入插值为稠密输入,Ikehata等^[27]引入自注意力机制来提高网络对稀疏输入的适应性。此外,Logothetis等^[28]优化CNN-PS结构,并设计先进的像素级训练策略,从而提高了表面法向量的估计精度。尽管上述方法一定程度上提高了CNN-PS对稀疏光源条件的适应性,但没有从本质上解决稀疏光源条件下精度下降的问题。

PS-FCN^[19]由特征提取部分、融合部分与法向量估计部分组成,其中特征提取部分利用权值共享的CNN结构提取不同光源图像中的邻域信息,融合部分采用最大池化操作融合不同光源下的图像特征,法向量估计部分利用融合特征高精度估计表面法向量。相比于CNN-PS结构,PS-FCN有三大优势:引入邻域信息更好地处理互反射、投射阴影等全局光照现象;少灯条件下法向量估计精度相对较高;法向量估计速度快。因此,本文认为PS-FCN结构更具有实际应用前景。为提高PS-FCN结构的精度,相关工作引入自注意力机制^[29]、朗伯体先验^[30]以及二元回归^[31]等结构,尽管上述方法均在一定程度上提高了PS-FCN结构的精度,但是这些方法均没有从逆向建模非朗伯反射角度改进网络结构,而有效的逆向反射模型是基于深度学习光度立体视觉方法成功的基础。

与CNN-PS^[20]结构相比,PS-FCN^[19]结构不仅在法向量估计部分体现了方向一致性线索,还在特征提取部分展现出与像素级逆向反射模型^[32]相似的特性。像素级逆向反射模型仅能有效逆向描述某种材料单个法向量的非朗伯反射行为,但是光度立体视觉网络需要有效逆向描述某种材料的全局非朗伯反射。Wang等^[33]根据各向同性非朗伯反射基本性质推导出共位光源逆向反射模型,该模型建立了某入射方向光线照射下观测像素值、共位光源照射下观测像素值、入射光线方向以及观测方向的点积与表面法向量以及入射光线方向点积之间的映射关系,通过引入共位光源照射下观测像素值有效提高像素级逆向反射模型的全局描述能力,其中共位光源是指距离相机很近的光源。Wang等^[33]根据该共位光源逆向反射模型设计深度学习网络,极大提高了表面法向量估计精度。尽管上述工作证明了共位光源逆向反射模型的有效性,但是该模型没有从各向同性反射基本定义出发进行推导,其逆向描述能力仍然存在理论上的缺陷。

为打破共位光源逆向反射模型的理论局限,本文设计了基于深度学习的逆向反射模型,包括三个阶段子网络,分别是方位角差子网络、逆向反射模型子网络与法向量估计子网络。第一阶段方位角差子网络旨在

以监督式学习方式提取与方位角差相关的图像特征,弥补共位光源逆向反射模型的理论不足;第二阶段逆向反射模型子网络以图像特征方式表征像素值到法向量以及入射光线方向点积之间的逆向映射关系;第三阶段表面法向量估计子网络充分利用前两个阶段子网络提取特征,实现表面法向量高精度估计。仿真结果证明,所提网络能够较好地同时处理100种典型各向同性非朗伯反射;实验结果表明,所提网络在标准Directional Lightings, objects of General reflectance and 'ground Truth' shapes (DiLiGenT)数据集上取得较高的法向量估计精度,平均法向量估计精度达 5.90° 。

2 基本方法

本节详细介绍基于深度学习的逆向反射模型,该模型基于平行光照射与正交相机投影两个基本假设进行推导,这是绝大多数非朗伯光度立体视觉研究的基本假设。

2.1 光度立体视觉技术基本原理

相机成像光度学模型^[34]是光度立体视觉技术的理论基础,早期光度立体视觉技术^[4]假设物体表面为朗伯反射,此时场景中某点的观测像素值可以表示为

$$o = I_{\text{int}} \mathbf{n}^T \mathbf{l}, \quad (1)$$

式中: o 表示观测像素值; \mathbf{n} 表示表面法向量; I_{int} 表示入射光线强度; \mathbf{l} 表示入射光线方向。 I_{int} 与 \mathbf{l} 均可通过标定获取。

当场景中存在 $m(m \geq 3)$ 个人射光线方向时,表面法向量的求解公式为

$$\mathbf{n} = \mathbf{L}^\dagger \tilde{\mathbf{o}}, \quad (2)$$

式中: $\tilde{\mathbf{o}} = [\tilde{o}_1, \tilde{o}_2, \dots, \tilde{o}_m]^T$, \tilde{o}_i 表示第 i 个人射光线方向对应的去除入射光线强度影响的像素值; $\mathbf{L} = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_m]^T$, \mathbf{l}_i 表示第 i 个人射光线方向,为 3×1 向量。

在真实场景中,物体表面多呈现复杂非朗伯反射,并且广泛存在阴影现象^[35],此时式(1)需要细化为

$$o = I_{\text{int}} \cdot s \cdot f_{\text{BRDF}}(\mathbf{l}, \mathbf{n}, \mathbf{v}) \cdot \max(\mathbf{n}^T \mathbf{l}, 0), \quad (3)$$

式中: $f_{\text{BRDF}}(\mathbf{l}, \mathbf{n}, \mathbf{v})$ 表示双向反射分布函数(BRDF); $\max(\mathbf{n}^T \mathbf{l}, 0)$ 表示附着阴影; s 表示投射阴影。当阴影出现时, $s = 0$;反之, $s = 1$ 。

非朗伯光度立体视觉技术根据式(3)实现表面法向量高精度估计,而式(3)的最大难点在于复杂非线性BRDF中包含未知表面法向量,本文设计基于深度学习的逆向反射模型来解决该问题。

2.2 共位光源逆向反射模型

共位光源逆向反射模型是在没有阴影的假设下推导的,此时场景中各向同性非朗伯表面某点的相机光度学模型^[32, 36]可以表示为

$$\tilde{o} = f_{\text{BRDF}}(\mathbf{n}^T \mathbf{l}, \mathbf{n}^T \mathbf{v}, \mathbf{l}^T \mathbf{v}) \cdot \mathbf{n}^T \mathbf{l} = f(\mathbf{n}^T \mathbf{l}, \mathbf{n}^T \mathbf{v}, \mathbf{l}^T \mathbf{v}). \quad (4)$$

在观测值 \tilde{o} 与变量 $n^T l$ 单调递增的假设下, 式(4)的反函数可表示为

$$n^T l = g(\tilde{o}, n^T v, l^T v), \quad (5)$$

但是该式中仍然存在与表面法向量相关的变量 $n^T v$ 。

为了处理变量 $n^T v$, Wang 等^[33]引入共位光源概念。共位光源是指与相机位置重合或者距离相机较近的光源。对于该光源照射场景中的每个点, 存在 $l = v$, 此时式(4)可简化为

$$o_0 = f_{\text{BRDF}}(n^T v) n^T v = f(n^T v), \quad (6)$$

式中: o_0 表示共位光源照射下的像素值, 且 o_0 仅跟变量 $n^T v$ 有关。

在 o_0 与 $n^T v$ 单调递增的假设下, 式(6)的反函数为

$$n^T v = g_0(o_0). \quad (7)$$

需要注意的是, f 单调性与 f_{BRDF} 单调性并不等价, f_{BRDF} 单调性是 f 单调性的充分不必要条件, 即当 f_{BRDF} 不单调时, f 也有可能单调。

将式(5)与式(7)联合起来, 可得到共位光源逆向反射模型, 即

$$n^T l = g(o, o_0, l^T v). \quad (8)$$

共位光源逆向反射模型的理论不足来源于式(4), 该公式由半参数化反射模型^[36]推导而来。各向同性反射半参数化模型可由 K 个非线性函数之和表示

$$f_{\text{BRDF}} = \sum_{k=1}^K \rho_k(n^T \alpha_k), \quad (9)$$

式中: ρ_k 表示第 k 个非线性函数; α_k 为 ρ_k 的参考方向, 是 ρ_k 集中分布的方向。在参考方向 α 位于入射光线方向 l 与观测方向 v 构成的平面上^[32, 36]的基本假设下, 推导出如式(4)所示的反射模型。

根据各向同性非朗伯反射的定义, 式(4)写为

$$\tilde{o} = f(n^T l, n^T v, \Delta\varphi). \quad (10)$$

式(10)与式(4)的区别在于变量 $\Delta\varphi$ 。 $\Delta\varphi$ 是入射光线方向 l 与观测方向 v 在法向量坐标系中的方位角之差, $\Delta\varphi$ 与变量 $l^T v$ 展示的角度在绝大多数情况下是不相同的。

幸运的是, 在共位光源条件下, $\Delta\varphi$ 为 0° , 这表明式(6)以及式(7)所示函数成立。此时, 理想的共位光源逆向反射模型应该写为

$$n^T l = g(o, o_0, \Delta\varphi). \quad (11)$$

共位光源逆向反射模型的理论不足在于利用变量 $l^T v$ 替换变量 $\Delta\varphi$, 但这一替换并不是等价的, 而是一种简化表示。此外, 变量 $\Delta\varphi$ 仍然与未知表面法向量有关, 不能利用光度立体视觉已知变量直接计算得到。通过设计合理的网络结构, 提取与变量 $\Delta\varphi$ 相关的图像特征, 弥补共位光源逆向反射模型的理论不足, 实现表面法向量高精度估计。

2.3 基于深度学习的逆向反射模型

根据式(11)设计了图1所示的基于深度学习的逆向反射模型, 包括三个子网络, 分别是方位角差子网络、逆向反射模型子网络与法向量估计子网络。第一阶段方位角差子网络提取与 $\Delta\varphi$ 相关的图像特征, 弥补共位光源逆向反射模型的不足; 第二阶段逆向反射模型子网络以图像特征方式表示逆向反射模型[式(11)], 实现表面法向量与入射光线方向点积的高精度估计; 第三阶段法向量估计子网络充分利用前两个阶段网络提取特征, 实现表面法向量估计, 形成端对端网络结构。

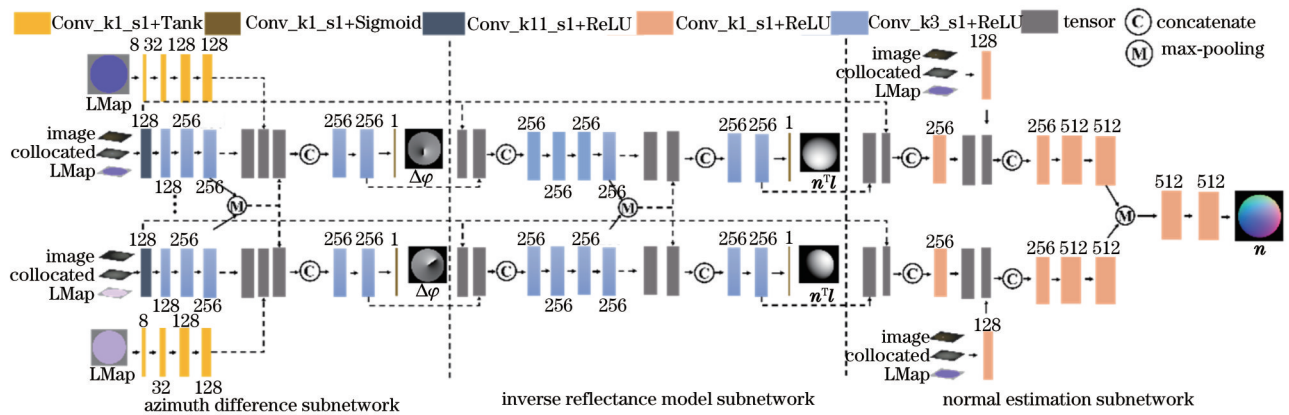


图1 基于深度学习的逆向反射模型

Fig. 1 Inverse reflectance model based on deep learning

2.3.1 方位角差子网络

第一阶段网络试图包含任意光源下像素值 o 、共位光源像素值 o_0 以及入射光线方向 l 与 $\Delta\varphi$ 之间的映射关系, 但是理论上 $\Delta\varphi$ 并不能由这 3 个变量计算得到, 其与入射光线方向以及表面法向量有关。在变量

$\Delta\varphi$ 的计算过程中, 需要将相机坐标系下表面法向量 $n = [\cos\varphi \sin\theta, \sin\varphi \sin\theta, \cos\theta]^T$ 变化为 $[0, 0, 1]^T$, 常见的坐标变换^[18, 21]是先将表面法向量绕着 z 轴旋转 $-\varphi$, 再绕着 y 轴旋转 $-\theta$, 即

$$T = \begin{bmatrix} \cos \theta \cos \varphi & \cos \theta \sin \varphi & -\sin \theta \\ -\sin \varphi & \cos \varphi & 0 \\ \sin \theta \cos \varphi & \sin \theta \sin \varphi & \cos \theta \end{bmatrix}. \quad (12)$$

在正交相机假设下,相机坐标系下观测方向固定为 $[0, 0, 1]^T$, 坐标变换后的观测方向为 $[-\sin \theta, 0, \cos \theta]^T$, 其对应的方位角固定为 π 。尽管上述变换还存在一个旋转不确定度,但是该旋转不影响 $\Delta\varphi$ 的数值。上述分析表明, $\Delta\varphi$ 仅与表面法向量以及入射光线方向有关,而与观测方向无关。

第一阶段子网络的输入端已经包含入射光线方向信息,有效提取与 $\Delta\varphi$ 相关特征信息的关键是引入与表面法向量相关的特征。受 PS-FCN^[19] 结构的启发,最大池化融合特征能够直接实现表面法向量估计,这表明该融合特征包含表面法向量信息,如图 1 左图所示。第一阶段网络结构包含两个输入分支与一个输出分支,以入射光线方向图 LMap 为输入的分支包含了 4 层 1×1 卷积,该分支激活函数设定为 Tanh,保证特征响应值在 $[-1, 1]$ 区间;以 o, o_0 以及入射光线方向 l 为输入的分支利用一层 11×11 卷积与三层 3×3 卷积提取图像邻域特征,最大池化融合特征根据该分支提取特征获得;输出分支连接两个分支提取特征与融合特征,实现对 $\Delta\varphi$ 的估计,将网络输出层激活函数设置为 Sigmoid,用于保证输出在 $[0, 1]$ 区间,训练过程中通过 $\Delta\varphi/(2\pi)$ 实现归一化。

2.3.2 逆向反射模型子网络

逆向反射模型子网络如图 1 中间图所示,将第一阶段子网络初始特征与终端特征联合起来作为输入,其中初始特征包含 o 与 o_0 信息,终端特征包含 $\Delta\varphi$ 特征,该输入以图像特征的形式展示了逆向反射模型[式(11)]。在该子网络中,融合特征根据 4 层 3×3 卷积层提取特征,并通过最大池化操作获得,该融合特征用于实现法向量与入射光线方向点积 $n^T l$ 的高精度估计。需要注意的是,该子网络的输出结果是 $s \cdot \max(n^T l, 0)$, 即该网络同时输出阴影信息与 $n^T l$ 。为了有效输出阴影信息,第二阶段网络将最大池化融合特征融入输出端,提高模型输出精度。阴影信息的最大特点是暗,即像素值低,在不同入射光线方向的比较中像素值低,在大片邻域区域像素值低,可根据该特点提取与阴影信息相关的图像特征。融合特征为邻域范围内不同位置光照下的最大特征响应值,很大程度上消除了阴影信息,将某光源照射下图像特征与该特征联合,可同时实现两个维度的比较,即邻域区域的比较以及不同光线方向的比较,从而有效输出阴影信息。训练过程中,阴影真值信息可通过阈值分割求得。

2.3.3 法向量估计子网络

根据第二阶段网络的输出,利用阈值方法去除阴影,即可实现表面法向量估计,但是阴影信息与 $n^T l$ 的

预测精度以及阈值选择都会产生误差累积,从而导致法向量估计精度下降。为了消除误差累积,设计了图 1 右图所示的法向量估计子网络,该网络以第一阶段子网络初始特征、第二阶段子网络终端特征与从原始数据 o, o_0 以及 l 重新提取的图像特征作为输入,直接估计表面法向量,实现端对端网络结构。

3 实验验证

通过渲染仿真训练数据集训练所提模型,采用消融实验证明网络结构设计的有效性,并通过仿真实验与真实实验分析所提方法的综合性能。消融实验与仿真实验均在仿真测试数据集上进行,该测试数据集包括 4 个物体,分别是 armadillo、bunny、dragon 与 sphere,其中前 3 个物体来自 Stanford 3D Scanning Repository 数据集^[37]。对于每个物体,在 1 个共位光源和 100 个均匀分布光源照射的条件下,渲染均一反射图像与 Spatially-Varying BRDF (SVBRDF) 图像,反射数据均来自 MERL 真实数据集^[21],图像尺寸设定为 256×256 ,图 2 展示了仿真测试数据集代表性图像。真实实验在 DiLiGenT 标准真实测试数据集上进行,该数据集是光度立体视觉领域常用的测试数据集,包含多种真实反射行为,能够有效测试光度立体视觉技术的有效性。为证明所提方法的先进性,以 CH20^[19]、IK18^[20] 与 WJ20^[33] 为对比方法,其中 CH20^[19] 与 WJ20^[33] 利用本节所提训练数据集重新训练。

3.1 训练数据集与训练细节

3.1.1 训练数据集

PS-FCN 结构^[19] 的训练数据集(下文简称“原始数据集”)可用于基于深度学习的逆向反射模型训练,但是该训练数据集存在两个缺陷:一是缺少 SVBRDF;二是部分渲染图像存在严重的数据缺失。为消除这些缺陷对网络性能的影响,本节在原始数据集的基础上设计了补充训练数据集。在补充训练数据集中,采用 Blobby 数据库^[38] 中 10 个物体(下文简称“Blobby 物体”)、MIT CSAIL Textured Models 数据集^[39] 中 7 个细节物体以及 Stanford 3D Scanning Repository 数据集^[37] 中 happy buddha 物体进行渲染(下文将这 8 个物体称为“细节物体”),代表性图像如图 3 所示。为补充 SVBRDF 数据,利用 15 种材料分布图,从 MERL 数据库^[21] 中随机挑选反射数据进行设计,图 3 中间图像展示了 SVBRDF 渲染效果,右边图像展示了材料分布图。为有效保存渲染数据,在路径追踪渲染过程中将光源强度设定为 5, gamma 设定为 $1/2.2$, 图像格式为 8 位图。补充训练数据集共包含由 Blobby 物体生成的 51840 组场景数据与由细节物体生成的 41472 组场景数据集,每个场景由 1 幅共位光源照射下的图像与 64 幅任意光源照射下的图像组成。

3.1.2 训练细节

利用 torch 库编写网络结构,在一块 GTX1080Ti

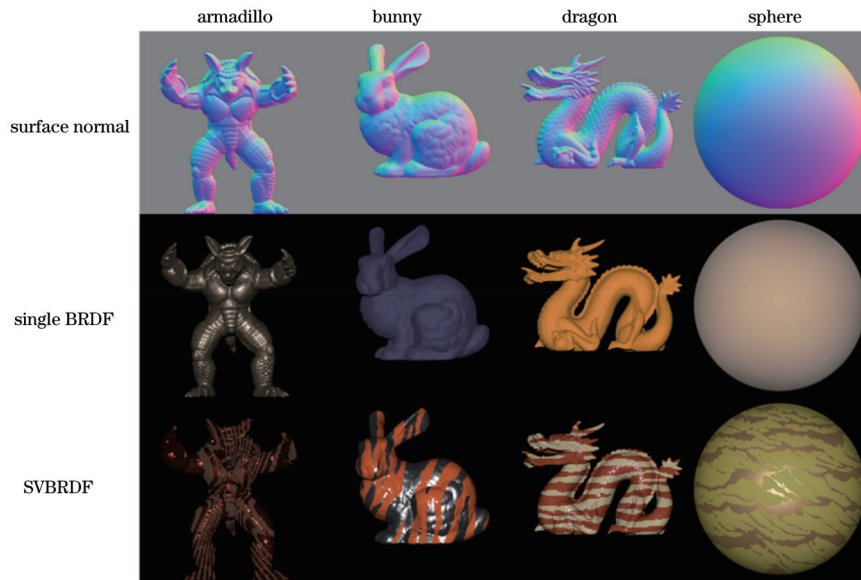


图 2 仿真测试数据集代表性图像

Fig. 2 Represented images of synthetic testing dataset

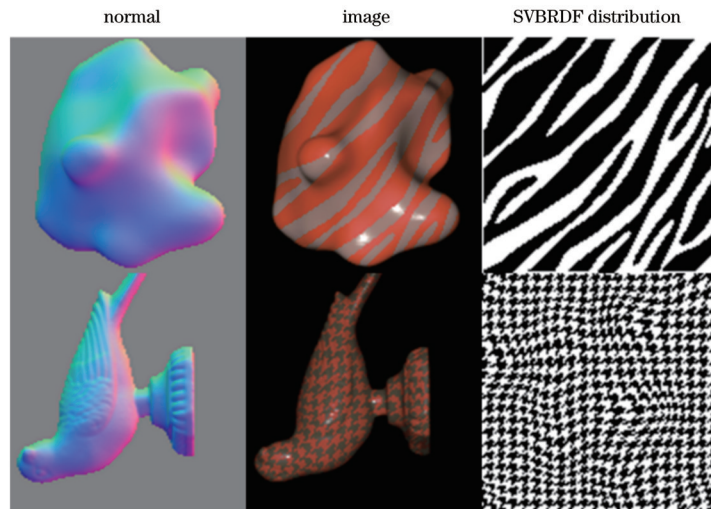


图 3 补充训练数据集代表性图像

Fig. 3 Represented images of proposed supplementary training dataset

显卡上训练基于深度学习的逆向反射模型。在训练过程中,所提补充训练数据集数据需要按照原始数据进行读取,同时对图像进行 gamma 反校正,消除入射光线强度的影响。所有图像需要变形并裁剪成 32×32 ,对图像每个通道进行颜色增强,其中补充训练数据集颜色增强范围设定为 $[1, 3]$,原始训练数据集颜色增强范围设定为 $[0.1, 3]$ 。此外,在训练与测试过程中,将场景平均像素值变化设为 $[10/255, 20/255]$,用于减小训练数据与真实测试数据之间的差异。

训练过程包括 3 个子网络分开训练阶段与联合训练阶段。在分开训练阶段中,随机误差范围设定为 $[-0.005, 0.005]$;在联合训练过程中,随机误差范围设定为 $[-0.01, 0.01]$ 。对于每一个子网络,批尺寸设定为 32,光源数量设定为 10,训练轮次设定为 40,其

中在第 26 个轮次与第 36 个轮次,学习率下降为原来的 $1/10$ 。此外,3 个子网络的损失函数均为均方根损失,损失函数分别表示为 $l_{\Delta\varphi}$ 、 l_{shading} 与 l_{normal} 。第一阶段子网络与第二阶段子网络的初始学习率设定为 0.0001,第三阶段子网络的初始学习率设定为 0.001。在联合训练过程中,学习率设定为 10^{-6} ,训练轮次设定为 5,批尺寸设定为 16,损失函数设定为 $l_{\text{normal}} + 0.1 \cdot l_{\Delta\varphi} + 0.1 \cdot l_{\text{shading}}$ 。表 1 详细记录了不同训练阶段超参数的设定情况。

3.2 消融实验

所提网络有两个重要特点:一是在第一阶段网络与第二阶段网络引入最大池化融合特征来助力网络提取有效特征;二是在第一阶段网络提取与 $\Delta\varphi$ 相关的图像特征,用于弥补共位光源逆向反射模型的不足,提

表 1 训练超参数设定

Table 1 Hyper-parameter settings of different training stages

Hyper-parameter setting	Stage 1	Stage 2	Stage 3	Combine
Initial learning rate	0.0001	0.0001	0.001	10^{-6}
Total epoch	40	40	40	5
Batch size	32	32	32	16
Learning rate decay epoch	[26,36]	[26,36]	[26,36]	-
Noise level	[-0.005,0.005]	[-0.005,0.005]	[-0.005,0.005]	[-0.01,0.01]

高 $n^T l$ 的映射精度。消融实验旨在证明这两个方面特点有效提高了网络性能,包括训练过程与测试过程。训练过程中,以新补充的Blobby数据集作为训练数据集,训练轮次设定为 30,批尺寸设定为 16,初始学习率设定为 0.0001,在第 26 个轮次学习率下降为原来的 0.1;测试过程在均一反射仿真测试数据集中展开。消融实验结果如表 2 所示,其中“-”表示比较误差不存在,数值表示平均绝对值误差(MAVE),该值越小表明精度越高。

表 2 所示的对比实验结果表明,通过引入最大池化融合特征,能够有效降低变量 $\Delta\varphi$ 与 $n^T l$ 的拟合误差,这证明引入融合特征有助于网络提取与变量 $\Delta\varphi$ 以及 $n^T l$ 相关的图像特征。 $n^T l$ 一列用于测试方位角差子网络的有效性,随着方位角差子网络的引入, $n^T l$ 的拟合精度提高了 17.5%。由该实验结果可知,相比于共位光源逆向反射模型,基于深度学习的逆向反射模型具有更高的逆向描述精度。

表 2 消融实验结果

Table 2 Ablation experiment results

Maxpooling	Stage 1	MAVE	
		$\Delta\varphi$	$n^T l$
×	×	0.101	0.064
✓	×	0.030	0.040
✓	✓	-	0.033

3.3 仿真实验

仿真实验测试了所提逆向反射模型对稠密光源场景、稀疏光源场景以及 SVBRDF 场景的法向量估计性能。

3.3.1 稠密光源场景实验结果

为测试所提逆向反射模型对稠密光源场景的性能,利用测试数据集中均一反射场景所有图像实现法向量估计,图 4 展示了 alum-bronze 反射 4 种算法的定性实验结果,可以看出,4 种算法都能在视觉效果上取得接近真值法向量的估计结果。

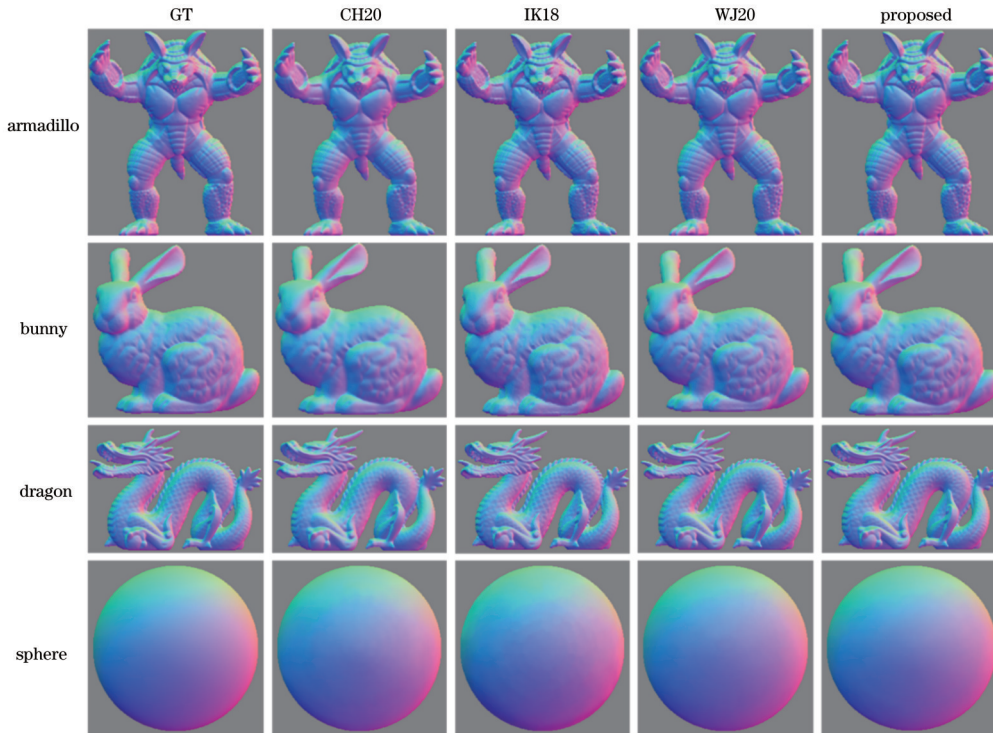


图 4 仿真实验定性结果

Fig. 4 Qualitative results of synthetic experiment

表 3 记录了定量实验测量的平均角度误差 (MAE)。MAE 是估计表面法向量与真值表面法向量之间夹角的平均值,是光度立体视觉领域常用的指标^[19],该指标越小,法向量估计精度越高。从表 3 可以看出,相比 WJ20^[33]、IK18^[20]与 CH20^[19] 3 种算法,所提方法在 4 个物体上均取得最小的法向量估计误差,平均误差分别下降了 24.1%、52.2% 与 54.2%,这证明所提方法对 MERL 数据集 100 种材料具有良好的泛化性。

表 3 稠密场景实验测量的平均角度误差
Table 3 MAE measured on dense light scene

Method	MAE / (°)				Average MAE / (°)
	Armadillo	Bunny	Dragon	Sphere	
Ours	2.3	1.9	2.2	2.3	2.2
WJ20 ^[33]	3.2	2.6	3.1	2.8	2.9
IK18 ^[20]	5.1	4.1	5.6	3.6	4.6
CH20 ^[19]	5.7	4.2	5.4	3.9	4.8

3.3.2 稀疏光源场景实验结果

减少光源数量不仅能够有效减小光度立体视觉装置的规模,还能有效提高光度立体视觉算法的效率。为测试所提方法对稀疏光源场景的有效性,利用仿真测试数据集中均一反射场景在不同稀疏光源下的图像估计表面法向量,实验结果如图 5 所示。可以看出,相比于 CH20^[19]与 WJ20^[33],所提方法取得最小的法向量估计误差,这证明所提模型能够有效处理稀疏光源场景。

3.3.3 SVBRDF 场景实验结果

在真实场景中,不仅存在均一反射表面,还广泛存在 SVBRDF 表面,而能否有效处理 SVBRDF 场景是光度立体视觉算法的重要衡量指标。为测试所提算法对 SVBRDF 场景的有效性,使用仿真测试数据集中 SVBRDF 场景所有图像实现表面法向量估计,实验结果如表 4 所示。相比于 WJ20^[33]、IK18^[20]与 CH20^[19] 3 种算法,所提方法仍然取得最高的表面法向量估计精度,表面法向量估计误差分别下降了 26.7%、52.2% 与 55.1%,证明所提方法能够有效处理 SVBRDF

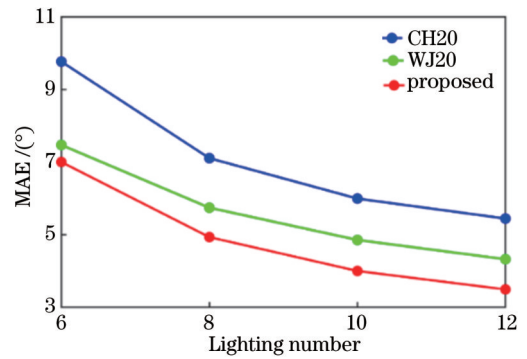


图 5 稀疏光源场景实验结果

Fig. 5 Experiment results on sparse light scene

表 4 SVBRDF 场景实验测量的平均角度误差
Table 4 MAE measured on SVBRDF scene

Method	MAE / (°)				Average MAE / (°)
	Armadillo	Bunny	Dragon	Sphere	
Ours	2.6	1.9	2.6	1.8	2.2
WJ20 ^[33]	3.4	2.5	3.5	2.6	3.0
IK18 ^[20]	5.1	3.8	5.8	3.6	4.6
CH20 ^[19]	5.9	4.0	5.6	3.9	4.9

场景。

3.4 真实实验

真实实验利用 DiLiGenT 标准真实数据集测试所提方法的有效性,在该实验过程中,以原始数据读取方式获取图像数据,并去除入射光线强度的影响。表 5 展示了每个阶段子网络的实验结果,图 6 定性展示了每个阶段子网络的实验结果。DiLiGenT 数据集中无法有效获取绝对真值的阴影信息,因此本实验利用法向量估计精度来间接反映 $n^T l$ 预测精度。表 5 所示的法向量估计误差从第一阶段到第三阶段逐渐递减,表明所提端对端网络能够有效提高表面法向量估计精度,三阶段网络结构设计合理;第四行法向量估计精度最高,表明所提联合训练策略能够有效提高网络性能。图 6 从左到右分别展示了第一阶段子网络、第二阶段子网络与第三阶段子网络实验结果,定性表明所提方法每个阶段网络的有效性。

表 5 每个阶段子网络实验结果

Table 5 Experiment results of every stage subnetwork

Object	Ball	Bear	Budd.	Cat	Cow	Gobl.	Harv.	Pot1	Pot2	Read.	Aver.
MAVE in stage 1	0.011	0.020	0.023	0.020	0.025	0.022	0.046	0.020	0.022	0.030	0.024
MAE in stage 2 / (°)	1.7	4.2	6.7	4.1	5.5	7.2	12.8	5.2	6.0	10.3	6.37
MAE in stage 3 / (°)	2.0	4.0	6.3	4.0	5.0	6.6	11.7	4.9	5.8	9.9	6.02
MAE of combine / (°)	2.0	3.6	6.3	3.9	5.1	6.5	11.6	5.0	5.2	9.8	5.90

表 6 展示了使用全部图像信息的实验结果,可以看到,所提方法取得最佳的法向量估计精度,其在 5 个物体上取得最小法向量估计误差,在另外 5 个物体上取得第二小法向量估计误差,这证明所提方法能够有

效处理各向同性非朗伯反射。对于具有明显全局光照效果的物体 Budd.、Harv. 和 Read.,所提模型的估计误差最小,这表明所提方法具有一定处理全局光照的能力。图 7 定性展示了 3 个物体法向量估计图与法向

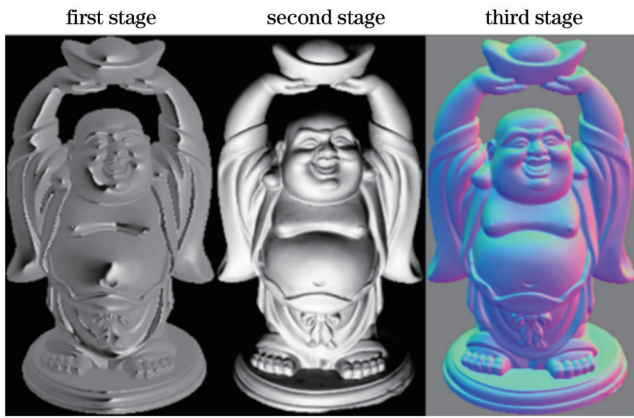


图 6 各阶段子网络定性实验结果

Fig. 6 Qualitative experiment results of every stage subnetwork

量估计误差图,可以明显看出所提算法的全局误差分布最小。

利用 10 个光源测试所提方法在少灯真实场景的实验效果,如表 7 所示,除 FI21^[28] 记录了 10 次实验平均法向量估计误差,其余算法均记录 100 次随机实验平均法向量估计误差。对于所提方法与 WJ20^[33] 方法,10 个光源是指 1 个共位光源与 9 个随意选择的光源。尽管 FI21^[28] 方法利用稀疏输入重新训练网络,但是该方法在少灯条件下的法向量估计精度仍然较差。从表 7 可以看到,所提逆向反射模型仍然取得最佳的表面法向量估计精度,其在 9 个物体上取得最小的法向量估计误差,证明所提方法能够有效处理稀疏光源场景。

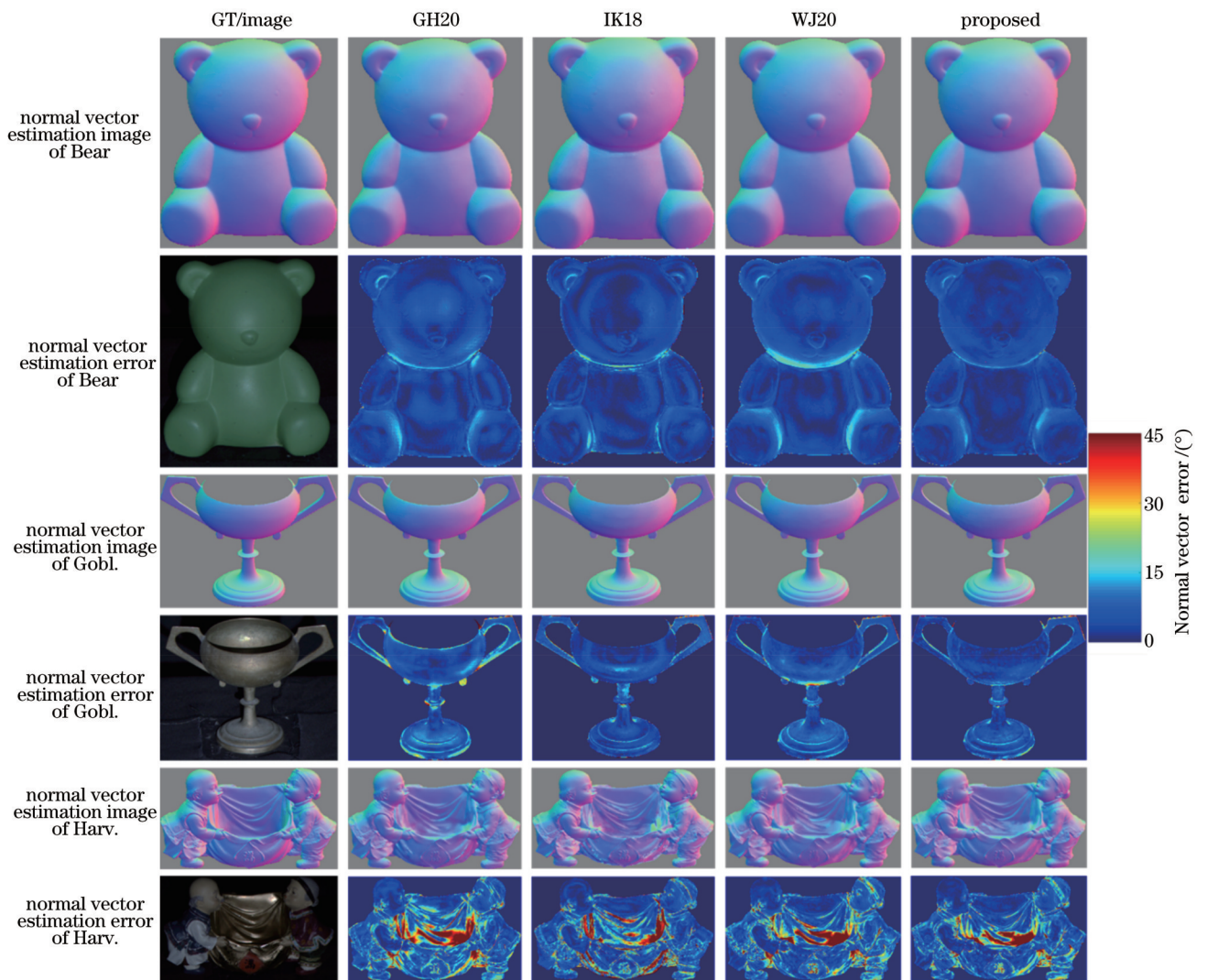


图 7 真实实验定性实验结果

Fig. 7 Qualitative results of real experiments

4 结 论

非朗伯反射问题是影响光度立体视觉技术性能的

核心问题,解决该问题的关键是构建有效的逆向反射模型。本文设计了基于深度学习的逆向反射模型,该模型可提升光度立体视觉技术处理各向同性非朗伯反

表 6 DiLiGenT 数据集全部光源条件下法向量比较结果

Table 6 Comparison results of normal vector on DiLiGenT dataset using all the images

Method	Normal vector estimation error / (°)										Average error / (°)
	Ball	Bear	Budd.	Cat	Cow	Gobl.	Harv.	Pot1	Pot2	Read.	
Ours	2.0	3.6	6.3	3.9	5.1	6.5	11.6	5.0	5.2	9.8	5.90
FI21 ^[28]	2.0	3.5	7.6	4.3	4.7	6.7	13.3	4.9	5.0	9.8	6.17
WJ20 ^[33]	1.6	4.6	6.9	4.7	5.2	7.8	13.0	5.6	6.6	10.2	6.62
IK18 ^[20]	2.2	4.1	7.9	4.6	8.0	7.3	14.0	5.4	6.0	12.6	7.21
CH20 ^[40]	2.4	4.6	7.5	4.8	8.0	8.9	13.9	5.7	6.7	10.5	7.30

表 7 DiLiGenT 数据集 10 个光源条件下法向量比较结果

Table 7 Comparison results of normal vector on DiLiGenT using 10 images

Method	Normal vector estimation error / (°)										Average error / (°)
	Ball	Bear	Budd.	Cat	Cow	Gobl.	Harv.	Pot1	Pot2	Read.	
Ours	2.1	4.3	7.2	5.2	6.2	8.3	13.8	6.1	6.7	10.7	7.06
WJ20 ^[33]	2.3	5.9	7.9	5.9	7.2	8.9	15.5	6.7	8.2	11.3	7.98
CH20 ^[19]	2.9	5.9	8.7	5.4	9.1	10.6	15.6	6.5	8.3	10.6	8.36
FI21 ^[28]	2.5	4.9	9.4	6.3	7.2	9.7	16.1	7.0	7.7	13.1	8.37

射的能力,实现表面法向量高精度估计。该模型具有 3 个特点:一是引入最大池化融合特征来表示表面法向量,有效提取与 $\Delta\varphi$ 相关的有效图像特征,弥补共位光源逆向反射模型的不足;二是以图像特征方式表征逆向反射模型,实现法向量相关变量 $\mathbf{n}^T \mathbf{l}$ 的高精度估计;三是阶段性训练策略,利用中间变量保证网络提取有效的图像特征,有效提高表面法向量估计精度。仿真实验证明,与已有逆向反射模型相比,所提基于深度学习的逆向反射模型能够更好地处理 100 种典型各向同性非朗伯反射。基于 DiLiGenT 标准数据集的真实实验证明,所提方法能够取得较高的法向量估计精度,平均法向量估计精度达 5.90° 。未来将致力于逆向建模更具挑战的各向异性非朗伯反射,并打破平行光照射与正交相机拍摄假设对光度立体视觉技术的束缚。

参 考 文 献

- [1] 朱可, 霍彦文, 武通海, 等. 基于光度立体视觉三维重构算法的微观磨损形貌原位测量原理及方法[J]. 机械工程学报, 2021, 57(10): 1-9.
Zhu K, Huo Y W, Wu T H, et al. Principle and method for *in-situ* measurement of micro-scale worn surface morphology based on 3D reconstruction with photometric stereo vision algorithm [J]. Journal of Mechanical Engineering, 2021, 57(10): 1-9.
- [2] 刘瑞鑫, 李立轻, 汪军, 等. 基于光度立体视觉的织物褶裥等级评定[J]. 东华大学学报(自然科学版), 2013, 39(1): 48-52, 59.
Liu R X, Li L Q, Wang J, et al. The evaluation of fabric pleat grade based on photometric stereo[J]. Journal of Donghua University (Natural Science), 2013, 39(1): 48-52, 59.
- [3] 洪海波, 万舒彪, 沈义平, 等. 基于光度-面结构光复合传感的大型航天构建机器人原位测量方法[J]. 航天制造技术, 2022, 5: 18-21.
Hong H B, Wan S B, Shen Y P, et al. Photometric and surface structure light composite sensing based robot *in-situ* measurement method for large aerospace component[J]. Aerospace Manufacturing Technology, 2022, 5: 18-21.
- [4] Woodham R J. Photometric method for determining surface orientation from multiple images[J]. Optical Engineering, 1980, 19(1): 139-144.
- [5] Hayakawa H. Photometric stereo under a light source with arbitrary motion[J]. Journal of the Optical Society of America A, 1994, 11(11): 3079-3089.
- [6] Kolagani N, Fox J S, Blidberg D R. Photometric stereo using point light sources[C]//Proceedings 1992 IEEE International Conference on Robotics and Automation, May 12-14, 1992, Nice, France. New York: IEEE Press, 2002: 1759-1764.
- [7] Shi B X, Mo Z P, Wu Z, et al. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo [J]. IEEE Transactions on Pattern Analysis, 2019, 41(2): 271-284.
- [8] Ikehata S, Wipf D, Matsushita Y, et al. Photometric stereo using sparse Bayesian regression for general diffuse surfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(9): 1816-1831.
- [9] Wu L, Ganesh A, Shi B X, et al. Robust photometric stereo via low-rank matrix completion and recovery[M]//Kimmel R, Klette R, Sugimoto A. Computer vision - ACCV 2010. Lecture notes in computer science. Heidelberg: Springer, 2011, 6494: 703-717.
- [10] Wu T P, Tang C K. Photometric stereo via expectation maximization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(3): 546-560.
- [11] Goldman D B, Curless B, Hertzmann A, et al. Shape and spatially-varying BRDFs from photometric stereo[J]. IEEE Transactions on Pattern Analysis, 2009, 32(6): 1060-1071.
- [12] Shi B X, Tan P, Matsushita Y, et al. Bi-polynomial modeling of low-frequency reflectances[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(6): 1078-1091.
- [13] Chung H S, Jia J Y. Efficient photometric stereo on glossy surfaces with wide specular lobes[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008.
- [14] Shen H L, Han T Q, Li C G. Efficient photometric stereo using kernel regression[J]. IEEE Transactions on Image Processing, 2017, 26(1): 439-451.
- [15] Higo T, Matsushita Y, Ikeuchi K. Consensus photometric stereo[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 2010: 1157-

- 1164.
- [16] Shi B X, Tan P, Matsushita Y, et al. Elevation angle from reflectance monotonicity: photometric stereo for general isotropic reflectances[M]//Fitzgibbon A, Lazebnik S, Perona P, et al. Computer vision - ECCV 2012. Lecture notes in computer science. Heidelberg: Springer, 2012, 7574: 455-468.
- [17] Hertzmann A, Seitz S M. Example-based photometric stereo: shape reconstruction with general, varying BRDFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8): 1254-1264.
- [18] Hui Z, Sankaranarayanan A C. Shape and spatially-varying reflectance estimation from virtual exemplars[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(10): 2060-2073.
- [19] Chen G Y, Han K, Shi B X, et al. Deep photometric stereo for non-lambertian surfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 129-142.
- [20] Ikehata S. CNN-PS: CNN-based photometric stereo for general non-convex surfaces[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision - ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11219: 3-19.
- [21] Matusik W, Pfister H, Brand M, et al. A data-driven reflectance model[J]. ACM Transactions on Graphics, 2003, 22(3): 759-769.
- [22] Enomoto K, Waechter M, Okura F, et al. Discrete search photometric stereo for fast and accurate shape estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4355-4367.
- [23] Santo H, Samejima M, Sugano Y, et al. Deep photometric stereo network[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2018: 501-509.
- [24] Santo H, Samejima M, Sugano Y, et al. Deep photometric stereo networks for determining surface normal and reflectances [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 114-128.
- [25] Li J X, Robles-Kelly A, You S D, et al. Learning to minify photometric stereo[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 7560-7568.
- [26] Zheng Q, Jia Y M, Shi B X, et al. SPLINE-net: sparse photometric stereo through lighting interpolation and normal estimation networks[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 8548-8557.
- [27] Ikehata S. PS-transformer: learning sparse photometric stereo network using self-attention mechanism[C]//32nd British Machine Vision Conference 2021, , November 22-25, 2021, Online. London: BMVC Press, 2021.
- [28] Logothetis F, Budvytis I, Mecca R, et al. PX-NET: simple and efficient pixel-wise training of photometric stereo networks[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 12737-12746.
- [29] Liu H Y, Yan Y H, Song K C, et al. SPS-net: self-attention photometric stereo network[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 5006213.
- [30] Ju Y K, Jian M W, Guo S X, et al. Incorporating lambertian priors into surface normals measurement[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 5012913.
- [31] Ju Y K, Dong J Y, Chen S. Recovering surface normal and arbitrary images: a dual regression network for photometric stereo[J]. IEEE Transactions on Image Processing, 2021, 30: 3676-3690.
- [32] Ikehata S, Aizawa K. Photometric stereo using constrained bivariate regression for general isotropic surfaces[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 2187-2194.
- [33] Wang X, Jian Z X, Ren M J. Non-lambertian photometric stereo network based on inverse reflectance model with collocated light[J]. IEEE Transactions on Image Processing, 2020, 29: 6032-6042.
- [34] 程岳. 基于光度立体的高质量表面重建研究[D]. 杭州: 浙江大学, 2013.
Cheng Y. High quality surface reconstruction using photometric stereo[D]. Hangzhou: Zhejiang University, 2013.
- [35] Argyriou V, Petrou M. Recursive photometric stereo when multiple shadows and highlights are present[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008.
- [36] Chandraker M, Ramamoorthi R. What an image reveals about material reflectance[C]//2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2012: 1076-1083.
- [37] StanfordThe 3D scanning repository[EB/OL]. [2023-02-02]. <http://graphics.stanford.edu/data/3Dscanrep/>.
- [38] Johnson M K, Adelson E H. Shape estimation in natural illumination[C]//CVPR, June 20-25, 2011, Colorado Springs, CO, USA. New York: IEEE Press, 2011: 2553-2560.
- [39] MIT CSAIL textured models database[EB/OL]. [2023-02-02]. <http://people.csail.mit.edu/tmertens/texttransfer/data/>.
- [40] Chen G Y, Han K, Wong K Y K. PS-FCN: a flexible learning framework for photometric stereo[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision - ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11213: 3-19.

Inverse Reflectance Model Based on Deep Learning

Wang Xi, Jian Zhenxiong, Ren Mingjun*

*State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao
Tong University, Shanghai 200240, China*

Abstract

Objective To enhance the capability of photometric stereo to handle the isotropic non-Lambertian reflectance, an inverse

reflectance model based on deep learning is proposed to achieve highly accurate surface normal estimation in this paper. Non-Lambertian reflectance is an important factor affecting the performance of optical measurements like fringe projection. To our best knowledge, photometric stereo is only one technology that could solve the effect of non-Lambertian reflectance in theory. Traditional non-Lambertian photometric stereo methods employ robust estimation, parameterized reflectance model, and general reflectance property to handle the non-Lambertian reflectance, which in essence adopts different mathematical technologies to handle the reflectance model. With the introduction of deep learning technology, it is possible to directly establish the inverse reflectance model, and the capability of photometric stereo to handle the non-Lambertian reflectance significantly increases. The represented supervised deep learning methods are CNN-PS and PS-FCN. The CNN-PS directly maps the observation map recording the intensities under different lightings to the surface normal according to the orientation consistency cue. The performance of this network significantly decreases if there are a small number of lights. PS-FCN simulates the normal estimation process of the pixel-wise inverse reflectance model and employs the neighborhood information to give a robust surface normal estimation for the scene with sparse light. The pixel-wise inverse reflectance model could not globally describe the non-Lambertian reflectance, which is supplemented by introducing collocated light recently. However, there still exist theoretical limitations in the collocated light-based inverse reflectance model. Therefore, this paper attempts to complete the theoretical defect of the collocated light-based inverse reflectance model by effectively extracting the image feature related to azimuth difference and designing the deep-learning-based inverse reflectance model.

Methods We first analyze the theoretical limitation of the collocated-light-based inverse reflectance model, then design the three-stage subnetworks of the proposed deep learning-based inverse reflectance model, and train the model by the new training strategies. The theoretical defect mainly comes from the assumption of Eq. (4), or in other words, the main direction α should lie on the plane extended by the l and v . Now, the BRDF input value $\Delta\varphi$ is simplified by the value $l^T v$. However, $l^T v$ is not identical to the $\Delta\varphi$ in most circumstances, and $\Delta\varphi$ is highly related to the unknown surface normal. The proposed inverse reflectance model based on deep learning is designed as shown in Fig. 1, which consists of three subnetworks, i. e., the azimuth difference subnetwork, the inverse reflectance model subnetwork, and the surface normal estimation subnetwork. The first-stage subnetwork attempts to map the image o under arbitrary lighting, the collocated image o_0 , and the lighting map l to the $\Delta\varphi$ map, and the max-pooling fused feature is introduced to represent the surface normal. The second-stage subnetwork achieves the ideal inverse reflectance model in an image feature way. The output of this subnetwork could be directly utilized to calculate the surface normal by the least-square algorithm, but the shadow thresholding value directly and dramatically influences the estimation accuracy. Thus, the third-stage subnetwork is designed to avoid error accumulation and achieve accurate surface normal estimation. To train the proposed network, the new supplement training dataset is designed to save the low-reflectance data and provide the SVBRDF scene. The three subnetworks are firstly trained separately to obtain the initial model parameters of every subnetwork and then combined to finetune the parameters.

Results and Discussions In this paper, the ablation experiment is utilized to prove the effectiveness of the network design, and the synthetic experiment and real experiment are adopted to analyze the performance of the proposed method. The PS-FCN, CNN-PS, and the network proposed by Wang *et al.*, denoted by CH20, IK18, and WJ20, are adopted as comparison methods in this paper. As shown in Table 2, the ablation experiment illustrates that the introduction of the max-pooling fusion feature benefits the extraction of the image features related to the $\Delta\varphi$ and the shading, and the azimuth difference subnetwork could effectively supplement the defect of the collocated light-based inverse reflectance model to better handle the isotropic reflectance. The synthetic experiment validates that the proposed method could achieve the best performance on the scene with dense lights, sparse lights, and SVBRDF. Figure 5 exhibits the superior performance of the proposed method on the sparse light scene compared with the WJ20, which shows the necessity of breaking the theoretical limitation of the collocated light-based inverse reflectance model. The real experiment based on the benchmark DiLiGenT dataset proves the state-of-the-art performance of the proposed method. Table 6 and Table 7 demonstrate that our method could achieve an average surface normal estimation accuracy of 5.90° for the real scene, and the performance of the proposed method significantly increases under the sparse light scene.

Conclusions We design the inverse reflectance model based on deep learning to handle the isotropic non-Lambertian reflectance, which completes the theoretical defect of the collocated light-based inverse reflectance model by effectively extracting the image feature related to the azimuth difference. The proposed model contains three subnetworks: the azimuth difference subnetwork, the inverse reflectance model subnetwork, and the surface normal estimation subnetwork. The first two subnetworks achieve the inverse mapping between the intensity and the dot product of surface normal and lighting direction, and the third network fully employs the image features extracted by these two subnetworks to accurately estimate the surface normal. The proposed method contains three characteristics, i. e., the introduction of max-pooling

fusion feature to extract the feature related to $\Delta\varphi$, inverse reflectance model based on the image feature, and stage training strategy. The ablation experiment proves the rationality of the network design, and the synthetic experiments validate that the proposed method could simultaneously handle classical 100 isotropic reflectances. The real experiments based on benchmark DiLiGenT dataset illustrate that the proposed method could achieve accurate surface normal estimation with 5.90° . The synthetic and real experiments validate the state-of-the-art performance of the proposed method. In future work, we would like to inversely model the challenging anisotropic reflectance and to break the limitation of parallel lighting and orthogonal cameras for photometric stereo.

Key words optical computing; deep learning; non-Lambertian reflectance; photometric stereo