基于多尺度卷积网络的非朗伯光度立体视觉方法

陈明汉, 任明俊*, 肖高博, 王晰, 付琳

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

* E-mail: renmj@sjtu.edu.cn

收稿日期: 2019-04-28; 接受日期: 2019-09-16; 网络版发表日期: 2020-02-10

国家重点研发计划(编号: 2018YFB1306800)、国家自然科学基金(批准号: 51675456)和天津大学精密测量技术与仪器国家重点实验室开放基金资助项目

摘要 光度立体视觉作为一种精细三维测量技术广泛应用于三维重构、缺陷检测、生物医疗等领域, 但传统反射模型对材质真实物理特性的反映能力有限, 对于如光亮金属等具有非线性光反射特性的非朗伯表面适应性不佳, 极大地限制了该技术的进一步应用. 本文提出了一种基于深度学习多尺度卷积架构的光度立体视觉算法, 实现了对非透明材质表面在任意光照条件下的高精度法向量恢复. 算法在深度网络中设计了多尺度卷积结构, 小尺寸卷积核强化了模型在光度物理原理上的表达, 使得模型在细节预测上具备优势, 大尺寸卷积核鼓励深度网络利用邻域特征, 提升模型克服阴影和区分多种材质的能力. 为进一步处理任意光照条件的问题, 算法设计了对入射光照向量空间的多分辨三维极坐标划分方法, 在整合输入图像信息的同时, 充分发挥了多尺度卷积的效能. 实验结果表明, 多尺度卷积深度学习架构有效集成了光度原理和深度学习二者的优势, 在保留精细三维形貌恢复能力的同时, 极大提高了针对非朗伯表面的适应性, 为光度技术的广泛应用提供了有力的技术支持.

关键词 深度学习, 光度立体视觉, 多尺度卷积

1 引言

光度立体视觉方法在1980年由Woodham^[1]最早提出. 该方法采集同一场景在不同光照条件下的多幅图片, 并对光反射和成像过程进行建模以构建图片像素值与被摄物体表面法向量间的函数关系, 进而求解物体表面法向量. 由于光度立体视觉方法能够在低成本条件下对物体表面进行高分辨率重建, 所以在精细重建、缺陷检测、生物医疗等领域有广泛的应用, 比如织物褶皱重建^[2]、金属表面检测^[3]、蔬菜秧苗叶片形

态测量^[4]、中医舌体的表面重建^[5]等.

鉴于反射模型的复杂性, 早期的光度立体视觉只能在表面朗伯体假设的简化下进行探索. 在朗伯体假设下, 高光和阴影被归结为噪声范畴, 基于此, 许多优化性研究工作得以展开. 比如Hayakawa^[6]提出了一种通过矩阵生长的策略来处理阴影, 找出计算矩阵中没有被噪声影响的数据来达到去除噪声影响的目的. Miyazaki等人^[7]利用随机抽样一致性RANSAC(random sampling consensus)方法, 并在表面以及色彩平滑的假设下, 自动将镜面反射所在的噪声成分排除在计算之

引用格式: 陈明汉, 任明俊, 肖高博, 等. 基于多尺度卷积网络的非朗伯光度立体视觉方法. 中国科学: 技术科学, 2020, 50: 323–334

Chen M H, Ren M J, Xiao G B, et al. Non-Lambertian photometric stereo vision algorithm based on a multi-scale convolution deep learning architecture (in Chinese). Sci Sin Tech, 2020, 50: 323–334, doi: 10.1360/SST-2019-0164

外, 吴仑等人^[8]通过鲁棒的主成分分析(robust principle component analysis, RPCA)去除图像噪声, 得到低秩矩阵和物体表面向量场, 然后再通过表面重建算法从向量场来恢复物体形状从而获得较高的法向量恢复精度。然而, 朗伯体假设对材质真实物理特性的反映能力有限, 因此上述研究对于现实非朗伯体材料反射特性的描述能力普遍不佳, 基于光度技术的表面重建精度尚有较大提升空间。

非朗伯体材料的反射特性以双向反射分布函数(bidirectional reflectance distribution function, BRDF)进行描述。在给定一种材质的条件下, 以光线入射方向和观测方向共4个角度参数作为变量, 就能够利用BRDF模型计算得到该表面对应的反射系数。BRDF模型相比朗伯体模型以更为普适的物理原理对反射过程进行描述, 因此在表面反射率的求解上具有更高的精度, 但同时也具有更高的计算复杂度。在BRDF模型基础上, 许多学者也开展了工作, Georgiades^[9]针对非朗伯体材质的表面恢复进行了研究, 在无需光源位置和强度或摄像机位置的信息的条件下, 准确地恢复物体表面的BRDF, 并应用于人脸重建, 取得了优秀的视觉效果。Goldman等人^[10]认为所有材料的BRDF可以由基本材料的BRDF组合得到, 并在单视图无参照物的情况下实现了未知材质BRDF的预测与重建。然而, 尽管BRDF模型对绝大多数材质的反射特性有良好的描述精度, 其仍旧存在一些问题: 一方面, 它依然是物理反射过程的一个近似描述, 模型能够达到的测量精度仍存在局限; 另一方面, BRDF模型较为复杂, 使得反射率的求解过程往往伴随大量优化迭代过程, 运算效率较低。

近年来, 深度学习方法被创新性地引入光度立体视觉领域, 以神经网络的前向传递过程取代了对传统反射模型的繁琐求解过程。深度学习方法在避免进行光度复杂建模与计算的同时, 还达到了引人注目的精度水平。深度网络以大量数据作为支撑, 其数据集由各类材料在多个光照下的表面采样图像和表面真实法向量图构成。通过训练网络, 即可得到对光度学映射关系的一个模型估计。深度学习方法一方面简化了传统光度的复杂模型, 在预测计算上避免了迭代步骤, 使运算效率相比传统依赖于迭代运算的优化算法有显著提升。另一方面, 深度网络模型不仅能够学习到光度方法的物理原理, 同时能够从训练数据中学习到一定的

先验知识, 这对于阴影、高光、异常点的处理具有优势, 相比传统方法鲁棒性更好, 近期的诸多研究表明其取得的预测精度能够达到更高。

DPSN(deep photometric stereo network)^[11], 它是深度学习应用于光度领域的开山之作。DPSN以固定的96个光照方向下的像素值作为输入, 通过dropout模拟的阴影层以及7个全链接层, 最终得到本像素点对应法向量的三通道输出。DPSN使用简洁的网络结构就达到了优秀的精度水平, 但是它需要限制入射光照方向为固定值, 这使得在应用上多有不便, 可优化的空间较大。

在DPSN之后, 近期Chen等人^[12]提出了基于全卷积神经网络实现的PS-FCN架构, PS-FCN创新性地将图像RGB三通道与对应三维光照方向向量XYZ整合为6通道张量作为输入, 这使得网络能够适应任意光照方向下的图像输入。同时, 网络中使用了邻域卷积和最大值融合方法处理输入特征, 使得它在克服光照阴影问题上有很大优势。但是PS-FCN大量使用3×3图像卷积进行邻域像素组合, 而并非同一位置不同图片上的像素进行组合, 这使得光度的基本原理某种意义上不能得到充分的体现。实验表明PS-FCN在部分物体的预测上会丢失细节信息, 光度方法所具备的高分辨率特征有所损失。

同一时期Ikehata^[13]提出了CNN-PS网络, CNN-PS同样是一个非常优秀的网络架构。其主要创新点在于对多光照方向输入图片的处理上, 网络结构本身则是在现有的经典深度学习网络DenseNet^[14]基础下构建的变体。CNN-PS将光照方向投射到xy平面, 对平面空间进行32×32的栅格化剖分, 并按照栅格次序逐个提取对应图片的像素, 构成32×32的观测图片作为网络输入。这种构建方式巧妙地将光照方向信息作为图片排序的依据, 最终实现了很高的计算精度, 同时非凸曲面的法向量预测上存在优势。但是CNN-PS在各项异性材质上表现的效果不佳, 例如铬钢和碳化钨。另外, CNN-PS的架构特点主要适应于光源较多的情况下, 因为它本质上需要依靠光照向量空间的邻域信息进行表面法向量回归, 而非图片像素的邻域信息。

本文在总结近期研究的基础上, 提出了一种基于深度学习的全新架构来描述光度立体视觉模型。网络首先采用多分辨空间极坐标对光照方向向量所在单位半球面进行划分, 有序化多图输入, 一方面使得网络能

够处理来自任意光照方向下的输入, 相比DPSN具备更好的应用能力; 另一方面, 基于光度理论设计的多分辨极坐标空间划分方式, 对不同方向的输入进行了一定程度上的筛选利用, 相比于CNN-PS的栅格划分方式, 它在理论上更能够提升网络预测精度. 其次, 在网络结构的设计上, 提出多尺度卷积的思想, 1×1 卷积操作有效强化了光度学原理的表达, 3×3 卷积操作则增强了网络在克服光照阴影以及识别多种材质上的能力. 多尺度卷积网络与PS-FCN网络一样都使用了全卷积神经网络作为架构主体, 而PS-FCN仅使用了 3×3 卷积, 这使得它在部分物体的预测上会丢失细节信息, 反观多尺度网络增加了 1×1 卷积, 则使得细节信息得以保留. 根据MERL(mitsubishi electric research laboratories)数据库^[15]仿真实验的结果, 本文证明了多尺度卷积深度网络具有较高的精度水平以及良好的计算效率. 同时本文设计了真实场景实验, 通过与其他深度学习架构的对比, 反映出了多尺度卷积结构在解决光度问题上的重要效能.

2 相关工作

2.1 光度仿真数据的构建

使用深度学习方法的前提是进行数据集构建. 在光度立体视觉中, 现有采集于现实环境的数据集数量较少, 并且采集代价较大, 因此主要以构建仿真数据集为主, 并以仿真数据集进行训练.

在传统光度立体视觉假设条件下, 以BRDF模型作为反射特性, 可以推导得到图像像素亮度 I_p 与物体表面法向量 $\mathbf{n}$ 之间的关系为^[16]

$$I_p = f_{\text{BRDF}}(\mathbf{l}, \mathbf{n}, \mathbf{v}) k_c I_s T \mathbf{n}^T \mathbf{l}, \quad (1)$$

其中, $\mathbf{l}$ 是入射光照单位方向向量, $\mathbf{v}$ 代表观测方向, 在远场条件下可以认为是常向量 $[0, 0, 1]^T$. f_{BRDF} 表示BRDF函数, 如图1所示, BRDF函数在定义时表示为自变量为 $\phi_i, \phi_e, \theta_i, \theta_e$ 4个角度参数的函数 $f_{\text{BRDF}}(\phi_i, \phi_e, \theta_i, \theta_e)$. 由于这4个角度参数可以由 $\mathbf{l}, \mathbf{n}, \mathbf{v}$ 确定得到, BRDF函数也可以写为 $f_{\text{BRDF}}(\mathbf{l}, \mathbf{n}, \mathbf{v})$, 因此在给定了材质和参数 $\mathbf{l}, \mathbf{n}, \mathbf{v}$ 之后, 即可唯一确定此表面点的BRDF反射特性. 式(1)中, k_c 是对成像模型的一个简化系数, I_s 是光源强度, T 是曝光时间. $k_c I_s T$ 可以合并认为是一个常数, 实际实验时, 需要通过标定的手段确定这个常数, 并使

图1 (网络版彩图) BRDF反射模型

Figure 1 (Color online) BRDF reflectance model.

用 $\tilde{T}_p = \frac{I_p}{k_c I_s T}$ 对图像进行预处理以去除其影响, 而在生成仿真数据时, 直接设置此常数为1即可, 因此式(1)成为

$$I_p = f_{\text{BRDF}}(\mathbf{l}, \mathbf{n}, \mathbf{v}) \mathbf{n}^T \mathbf{l}. \quad (2)$$

f_{BRDF} 函数来自各类材质数据集. 本文使用MERL数据集, 它提供了100种材料的BRDF数据. $\mathbf{n}$ 的获取也可以基于数据集, 比如DPSN使用的Blobby Shape Dataset数据集^[17]就提供了10种不同形状物体的表面法向量分布以及对应的掩模图片. 在得知 f_{BRDF} 和 $\mathbf{n}$ 的条件下, 通过给定 $\mathbf{l}$ 就能够生成对应光照下的仿真渲染图片. 由于物体的形状可能非常复杂, $\mathbf{n}^T \mathbf{l}$ 出现负值是常态, 此时对应表面上出现附着阴影(self-shadow)^[18]的情形. 然而仅仅依赖上述的计算无法模拟出投射阴影(cast-shadow)^[18]的效果(图2), 需要根据物体表面形状特征来额外添加. 最终实际使用的公式应当为

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

图2 (网络版彩图)附着阴影和投射阴影

Figure 2 (Color online) Attached shadow and cast shadow.

式中, $\max(0, \mathbf{n}^T \mathbf{l})$ 部分反映了附着阴影. 参数 s 的取值为 0 或 1, 体现为投射阴影.

2.2 光照分布的优化

根据光度原理, 特定方向范畴的光源光照会导致在光度求解过程中具有较大的条件数, 这对于获取高精度结果是不利的, 具体分析如下.

光度模型的最终求解式可以归纳为方程组:

$$\mathbf{L}^T \mathbf{n} = \mathbf{I}, \quad (4)$$

式中, $\mathbf{n}$ 为表面法向量, $\mathbf{L}$ 是 $3 \times m$ 矩阵 (m 表示入射光照数量), $\mathbf{I}$ 是 m 维亮度向量. 对于单个方程:

$$l_x n_x + l_y n_y + l_z n_z = i. \quad (5)$$

以图1的半球模型为例, 若光照方向 $\mathbf{l}$ 大量来源于半球顶端的位置, 也即式(5)中 l_x 和 l_y 接近于 0, 那么就会使得方程接近于 $l_z n_z = i$, 对 n_x 和 n_y 的预测难度加大. 反之若处于半球底端, 对 n_z 的预测难度就会加大, 因此光度立体视觉系统的设计往往避免在顶端和底端布置密集的光源, 因为中段位置的光源对于提升光度计算精度有较高的效益. 正是基于此理论, 本文设计了针对输入光照方向的多分辨极坐标划分方法, 充分利用中段位置的光照图像, 从理论上提升了模型的预测精度, 详见第3.1节所述.

2.3 多尺度卷积思想

多尺度融合的思想在各个领域内都有着广泛应用. 其实现形式有多种, 比如多模态的方法^[19]、多特征的方法^[20]以及多层次的方法^[21].

在深度学习中, 多尺度融合可以通过多尺度卷积方法实现, 并且已经在部分深度学习架构中得到了应用. 其中最为经典的就是 GoogleNet^[22] 的 Inception Module, 其将上一层的输入分别进行多尺度卷积核的作用, 在输出上将各个卷积核的输出沿特征通道方法叠合(concatenate)在一起, 整合作为下一层的输入, 有利于本层网络利用上一层的多尺度特征进行更好的预测.

需要注意的是, 本文所提出的方法与诸如 Inception Module 在内的其他多尺度卷积方法存在很大不同. 首先, 多尺度卷积结构在处理光度问题上存在特定的优势, 并且在本文所述的网络输入方式下, 多尺度卷积具备良好的可解释性, 并能够发挥出优秀的效

能. 其次, 本文所需要处理的是回归问题, 而诸如 GoogleNet 等其余网络主要处理的是分类问题. 分类问题依赖于反复卷积所得到的高阶特征进行判断, 因此 concatenate 操作是有效的, 但在回归问题中, 每一个像素都需要回归一个特征输出, 因此有必要维护像素级尺度的特征传递, 否则最终层得到的特征输入都是较大尺度范围内的结果, 这就会导致最终预测的结果显得模糊, 丢失细节信息. 从这点上看来, 对输出进行 concatenate 操作会导致单个像素尺度上的特征 (1×1 卷积结果) 随着网络层数的增加而不断减少, 使得模型丢失了高分辨细节预测能力. 因此本文提出了加权组合的方式以融合每层的输出, 维护了像素级尺度的特征传递, 详见第3.2节所述.

3 算法原理

3.1 空间有序化策略

深度网络应用于光度立体视觉上的一个困难就是如何同时处理多张图像的问题. 由于不同的图像是在不同的光照方向下采集的, 因此使用光照方向对图片进行一定的规则化排布是一个合理的切入点.

对于单位化光照方向向量 $[l_x, l_y, l_z]$, 存在约束 $l_x^2 + l_y^2 + l_z^2 = 1$, 因此可以仅使用两个变量对其进行描述, 那么对向量的划分势必也是二维化的. CNN-PS 提供了一种 xy 平面栅格划分的方法, 将光照方向向量投影到 $[0, 1]$ 范畴的 xy 平面上, 进行 32×32 的栅格划分, 从而建立了各个图片的二维排布顺序. 这种方法给出了一种输入图片的排布思路, 但是存在两个方面的问题.

(1) 划分的分辨率较高 (32×32), 导致对训练数据集的要求较高. 训练数据集需要提供大量且光照分布均匀的图像, 否则会造成有效栅格过于稀疏, 影响训练结果.

(2) xy 平面均匀划分未必是符合光度原理的较好选择. 如2.1节所述, 一般的光度系统在顶端和底端会布置较少的光源以避免求解过程中具有较大的条件数. 在这种情况下, 选用 xy 平面均匀栅格划分, 就会在中心和边缘处的栅格(分别对应顶端和底端位置)造成空间浪费. 显然可以考虑减少这两个位置的划分分辨率, 转而增加中段部分的分辨率, 毕竟中间部分的图片对于提升光度计算精度有较高的效益.

基于以上考虑, 本文提出一种基于空间极坐标描述的多分辨率半球划分方法, 如图3所示, 空间极坐标由仰角 θ , 方位角 ϕ 以及半径 r 三个参数描述. 由于光照方向向量为单位化向量, 因此所在的半球域 $r=1$. 图4给出了划分方法的俯视视图, 取仰角划分位置为 $\theta=[10, 20, 33, 45, 58, 70, 90]$, 并分别对由此形成的7个圆域或圆环域在 ϕ 方向上均匀划分子域数量 $[1, 3, 11, 17, 15, 12, 5]$, 使得总和为64个子域. 根据前述的思想, 算法在接近顶端 $\theta=0^\circ\sim 20^\circ$ 以及底端 $\theta=70^\circ\sim 90^\circ$ 的位置采用低分辨率进行球面划分, 而中间部分则采用较高分辨率划分. 形成的64个子域自顶向下按照如图4所示的方式进行编号, 形成一定的空间顺序, 并进一步将各个子域上对应图片转化为灰度图, 按照对应的序号依次在通道维度上进行整合, 这样就得到了卷积网络的64维输入. 对于图片空缺的子域, 使用0值构成的灰度图片进行填充. 对于多个图片同时属于同一个子域的情况, 对图片间进行最大值融合后再填充, 这是为了尽可能地抵抗阴影的出现.

多分辨率极坐标划分方式基于光度学原理, 充分利用了光照方向的空间分布信息来整合多张图片. 本文按照64维输入的情形设计了划分分辨率, 降低了对数据集的要求, 对于单个样本不需要大量的图片供应. 同时, 通过这种划分方式, 从光度学原理上提升了空间各

图3 (网络版彩图)空间极坐标定义

Figure 3 (Color online) Definition of spatial polar coordinates.

个输入的利用效益, 并使网络适应任意光照方向与任意数量的图片输入, 方法简单而实用.

需要说明的是, 对输入图片进行空间排序是非常重要的过程. 一方面, 如果不进行排序, 那么光照方向信息就没有被网络所利用, 实验表明此时网络难以收敛或者仅能收敛到一个精度误差较大的最优点处. 另一方面, 在实际应用中, 空间排序划分的策略要依据实际光照分布的情况进行选择. 尽管选择上文所述的划分方式同样能得到一个精度较好的结果, 但是有可能会造成部分子空间位置上始终缺失光照图片, 信息

图4 (网络版彩图)极坐标半球划分

Figure 4 (Color online) Polar hemisphere division.

得不到充分利用. 此时改变划分方式, 选择一个更适合的划分手段, 对于获取更高的测量精度是非常有意义的.

3.2 多尺度卷积结构

在卷积神经网络中, 使用高尺寸卷积核(如 3×3)能够利用图片像素的邻域信息. 从光度问题的角度来看, 这对于克服阴影, 区分多种材料以及提升运算效率都存在优势. 但就本文的观点, 大量地使用高尺寸卷积核, 难以保证网络能够真切学习到光度的物理原理. 一个可能的结果是网络更多地依赖于训练集的邻域先验知识做出估计, 使得网络难以具备光度的高分辨特征, 在细节丰富表面上的表现效果不佳. 同时本文注意到, DPSN和CNN-PS都使用了逐个像素预测的计算方式, 输入的信息是同一个像素在不同光照条件下的亮度结果. 这种结构强化了光度原理的表达, 但是由于是逐个像素预测, 它们忽略了像素邻域信息, 对于阴影和高光的抵抗能力不及PS-FCN. 那么, 一个自然的想法就是设计一种网络结构, 既能利用像素邻域信息, 又能够强化光度原理表达. 基于这样的目标, 本文提出了多尺度卷积结构来处理光度问题, 能够在利用邻域信息的同时又强化网络对光度基本原理的学习.

多尺度卷积(multi-conv)的基本结构如图5所示. 依照第3.1节图片有序化的结果, 网络输入在通道维度上

被堆叠形成64维输入. 这种图片堆叠的方式, 使得模型能够利用 1×1 卷积直接实现DPSN网络的处理结构, 也即等价于同一像素在不同光照下结果经由全链接层得到输出. 同时, 网络能够使用 3×3 卷积提取并利用像素的邻域特征, 提升对材质的辨识能力以及对阴影的克服能力.

简单而言, 多尺度卷积结构相当于对上一层的输出分别进行 1×1 卷积和 3×3 卷积, 得到的两个结果 $\text{Conv}_{1 \times 1}$ 和 $\text{Conv}_{3 \times 3}$ 并进行组合最终得到本层的输出.

由于光度问题是一个回归问题, 对于每一个像素都需要预测一个值, 如果合并方法采用GoogleNet的concatenate方式会逐层丢失像素级尺度的特征, 因而算法采用了加权组合的方法, 这样就避免了像素级尺度的特征的逐层减少, 也即

$$\text{out} = w_{1 \times 1} \text{Conv}_{1 \times 1} + w_{3 \times 3} \text{Conv}_{3 \times 3}. \quad (6)$$

需要注意的是, 两类卷积结果都具备有益的信息, 难以人为衡量两者的比例, 因此令组合权重 $w_{1 \times 1}$ 和 $w_{3 \times 3}$ 由网络自主学习得到, 也参与网络的训练, 并添加约束:

$$w_{1 \times 1}, w_{3 \times 3} \geq 0, \quad (7)$$

$$w_{1 \times 1} + w_{3 \times 3} = 1. \quad (8)$$

对于每一个多尺度卷积层, 都存在两个权重参数, 权重参数并不是层间共用的.

图5 (网络版彩图)多尺度卷积结构

Figure 5 (Color online) Structure of multi-scale convolution.

3.3 模型构建与训练

网络模型的基本结构如图6所示, 主体是全卷积神经网络, 并在其中增加了多尺度卷积结构.

主干由多尺度卷积层构成, 并增加一定的残差连接以支持网络加深, 最终形成的三通道结果进行归一化得到各像素上预测的法向量值.

网络中增加了两个下采样支路以增强感知域, 从而强化邻域特征的利用. 由于下采样部分的引入目的在于更好地利用邻域特征, 因此使用 3×3 卷积和Max pool实现, 而不必使用 1×1 卷积.

由于存在两个反卷积(deconvolution)上采样(图6中的Deconv), 为避免上采样引入的精度损失, 将两个上采样的结果分别引出支路进行损失定义. 如果预测正确, 那么进行简单的维度缩减后, 应当能够得到一个真实法向量图的下采样结果. 通过这种方法, 分别对两个下采样的结果添加损失, 从而降低上采样引入的误差.

损失函数使用预测法向量和真实法向量之间的平方误差之和:

$$\text{loss} = \sum_{\text{mask}(i,j)>0} (N_{ij}^{\text{pred}} - N_{ij}^{\text{gt}})^2, \quad (9)$$

式中, N_{ij}^{pred} 代表像素 (i, j) 上得到的预测法向量, N_{ij}^{gt} 则代表像素 (i, j) 上的真实法向量, mask是场景对应的掩

模图片, 表征像素 (i, j) 是前景还是背景.

需要注意的是, 对于1/4和1/2的下采样支路同样设置平方误差损失, 只是此时作为对比的真实法向量分别是1/4和1/2的采样结果.

3.4 算法实现

实验使用PS-FCN论文所提供的数据, 来源于blobby shape dataset^[17]以及the sculpture shape dataset^[23]在Mitsuba^[24]软件下根据光度原理的渲染结果. 为提升训练效率, 将所提供的原始图片数据预先进行空间有序化处理, 并在二进制编码压缩后存储为tfrecord文件以进行快速处理. 网络基于深度学习框架tensorflow进行构建, 并以tfrecord文件作为输入执行多线程训练. 运行电脑CPU为Core i7-4790K 4.00 GHz, 内存RAM为16 GB, GPU为GeForce GTX 1050 Ti. 学习率初始值为 7×10^{-4} , 以指数形式每隔500次迭代进行衰减, 衰减系数为0.01, 共训练30个epoch, 时间总消耗18 h.

4 实验设计与效能验证

为验证模型的效能, 本文设计了两个测试实验: 一则是在MERL标准数据库上进行了仿真数据的测试实验; 另一则是搭建了实际装置进行真实场景的测试实验.

图6 (网络版彩图)网络结构图

Figure 6 (Color online) Network structure.

4.1 标准数据库仿真实验

为验证模型的效能, 以bunny模型作为测试模型. 实验选取了40个在单位半球空间上均匀分布的光源位置, 依照2.1节所述渲染原理, 渲染得到MERL数据集下100种材质的渲染图片. 图7给出了光源的分布状况以及在PVC材质下的渲染示例(为便于观察, 图片亮度进行了一定放大), 分辨率为256×256.

渲染得到的图片具备RGB三个通道, 首先将其转换为灰度图. 然后根据3.1节所述的原理, 依照每张图片对应光照方向向量在空间的分布位置, 对图片进行排序, 形成64维输入, 输送给网络进行法向量估计. 图8给出了网络输出的一个例子, 并给出了估计法向

图7 (网络版彩图)光源分布和渲染示例

Figure 7 (Color online) Light source distribution and rendering examples.

图8 (网络版彩图)预测结果(alum-bronze)示例

Figure 8 (Color online) Prediction result (alum-bronze).

量与真实法向量之间的角度误差分布图.

对于每一个像素, 模型都能够对其所在物体表面单位法向量给出一个估计值 $\hat{\mathbf{n}}$, 估计值 $\hat{\mathbf{n}}$ 与真实值 $\mathbf{n}$ 之间的偏差往往以余弦相似度的形式进行描述:

$$\text{similarity} = \cos(\theta) = \hat{\mathbf{n}}^T \mathbf{n}, \quad (10)$$

式中, $\arccos(\hat{\mathbf{n}}^T \mathbf{n})$ 则反映了向量间夹角 θ 的大小, 角度值介于 $0^\circ \sim 180^\circ$ 之间, 作为估计向量和真实向量之间的角度偏差. 实验时对所有掩模有效的图片像素都可以计算一个角度偏差值, 并提取它们的均值作为模型在本材质上的估计偏差, 也即

$$\frac{1}{N} \sum_{i=1}^N \arccos(\hat{\mathbf{n}}_i^T \mathbf{n}_i). \quad (11)$$

为体现模型的效能, 与现有深度学习算法PS-FCN, CNN-PS(分为 $K=1$ 和 $K=10$ 两种情形), 以及传统光度学领域非常杰出的CBR算法(constrained bivariate regression)^[25], 在同样的40个上述光源下进行实验, 并记录了100种材料下的角度平均误差值与本文的模型进行对比, 结果参见图9和表1所示.

对各个算法除了在精度方面进行了对比, 在速度上同样也进行了对比. 统一运行环境: CPU为Core i7-4790K 4.00 GHz, 内存RAM为16 GB, GPU为GeForce GTX 1050 Ti, 以分辨率为256×256的图片作为输入. 根据记录, 实验表明4类算法在不同的材质表现上各有差异. 在精度上, CNN-PS($K=10$)表现得最好, 但是它需要运行 K 次求取均值, 并且逐像素预测的过程难以使用GPU加速, 造成其损耗的时间最大. PS-FCN具有最少的运行时间, 因为它参数量较少, 并且只需要通过全卷积网络一次便可以得到所有像素的结果, 并且全卷积的架构非常利于GPU并行加速, 但是其在精度上表现不佳. 相比之下, 本文的多尺度卷积网络既在精度上表现优异, 同时在运算速度上也具备一定优势, 具备巨大的实际工程应用潜力.

4.2 真实场景实验

为充分验证模型的实用性, 搭建了真实光度立体视觉实验平台以用于模型测试(图10). 该平台的上侧曲面上布置了47盏20 W功率的点光源, 顶端中心布置了一个相机. 其外部覆盖着不透光布料, 以实现小型物体在暗室环境下的光度测量. 整个平台拍摄距离约1.8 m, 相机型号为大恒相机水星系列MER-502-

图 9 (网络版彩图)平均角度误差对比

Figure 9 (Color online) Average angle error comparison.

表 1 MERL实验结果记录表

Table 1 MERL experiment result

算法	平均角度误差 (°)	平均运行时间 (s)
PS-FCN	7.147	0.124
CBR	4.397	6.779
CNN-PS(K=1)	5.354	5.846
CNN-PS(K=10)	3.819	45.043
本文方法	4.27	0.632

图 10 光度立体视觉实验平台

Figure 10 Experiment platform of photometric stereo vision.

79U3C工业相机, 实验时采集分辨率设定为 612×512 . 该平台控制软件基于Qt开发, 提供了友好的人机界面, 能够实现对相机采集和光源开关的手动控制以及

自动控制.

实验之前, 预先需要对场景的光照方向以及光照强度进行标定. 光照方向通过于场景中放入镀铬钢球进行标定, 光源强度通过放入标准白板标定^[16]. 标定得到的光照强度需要对采集图片进行预处理, 消除不同光源之间光强的差异. 标定得到的各光源光照方向则作为图片排序输入的依据.

实验选择陶制小型茶碗作为测试对象, 直径大小为50 mm, 高度为22 mm, 以倒置的方式进行拍摄. 如图11所示, 碗身部分为平滑曲面, 碗底部分存在毫米级尺寸的刻字纹理. 此测试对象能够分别对模型在平滑表面以及细节丰富表面上的预测效能进行考察. 拍摄得到的47张图片经过光照强度预处理后, 与光照方向一并输入模型. 实验分别在本文的多尺度卷积模型、PS-FCN以及CNN-PS上进行了测试, 并将得到的表面法向量预测结果进行对比.

测试结果如图12所示, 为便于观察, 将平滑位置以

图 11 (网络版彩图)实验对象

Figure 11 (Color online) Experimental object.

图 12 (网络版彩图)实验结果对比

Figure 12 (Color online) Comparison of experimental results.

及底部细节丰富的刻字区域进行了放大。

根据结果可以明显发现PS-FCN对于底部细节区域的恢复效果不好, 而CNN-FS和多尺度卷积网络都能够得到良好的效果。其原因主要是PS-FCN架构中大量的 3×3 卷积对于像素级特征的传递能力不足, 而CNN-FS由于逐像素的预测机制, 多尺度网络由于 1×1 卷积的作用, 使得像素级尺度的特征得以保持, 在面对纹理细节丰富区域的法向量预测问题时, 仍旧能够保持良好的预测精度。

另一方面, 在平滑区域上, 可以发现CNN-PS的结果存在部分噪点, 尤其是接近边缘处存在的问题。主要原因是CNN-PS进行逐像素预测, 没有依赖像素

邻域信息, 而PS-FCN和多尺度卷积网络都存在大量的 3×3 卷积提取像素邻域信息, 因此对于表面的平滑效果会更好。

真实实验的结果集中反映了多尺度卷积结构的优势所在。 1×1 卷积强化了光度原理的表达, 使得模型在细节丰富位置上具有良好表现。同时 3×3 卷积整合利用像素邻域特征, 避免了局部噪点, 使得模型在平滑区域上也能得到良好效果。真实实验表明多尺度卷积网络在实际工程应用上具备一定的潜力, 在细节丰富表面和平滑表面都能够有更好的精度表现。

5 结论与展望

本文提出了多尺度卷积深度光度立体视觉算法, 使用同一物体在不同光照方向下的采集图像, 对其表面法向量进行高精度预测。该算法在设计过程中结合了光度立体视觉和深度学习两种方法的优势, 在显著改善细节恢复质量的同时实现了对多种材料和复杂形状表面的高效率高精度法向量预测。从光度学角度, 设计了光照方向向量的空间极坐标划分策略, 并且在网络中增加了 1×1 卷积实现单一像素间特征的提取, 强化了网络在纹理细节丰富的表面上的预测能力。从深度学习角度, 在网络中增加了两个下采样支路以增大网络感知域, 并使用大量的 3×3 卷积层提取像素邻域信息, 强化了网络抵抗阴影与辨识多种材质的能力。根据仿真数据集和真实实验的结果, 多尺度卷积深度网络具备多个方面的优势: 第一, 算法在朗伯体材质和非朗伯材质上都能够取得很高的测量精度, 精度水平超越了多个传统经典算法; 第二, 多尺度卷积结构能够保持表面高分辨细节特征, 这是其他基于深度学习的光度模型难以企及的; 第三, 模型计算速度快, 避免了传统算法中大量的迭代过程, 应用前景较好。鉴于多尺度卷积深度光度立体视觉技术存在的多个优势, 其势必将对光度技术的广泛应用提供有力的技术支持。

参考文献

- Woodham R J. Photometric method for determining surface orientation from multiple images. *Opt Eng*, 1980, 19: 1-22
- Xu J M, Wang J. Fabric surface shape using method of 3D surface reconstruction based on photometric stereo images (in Chinese). *J Donghua Univ*, 2007, 33: 231-235 [徐建明, 汪军. 基于光度立体图像三维重建的起皱织物表面形态. *东华大学学报(自然科学版)*, 2007, 33: 231-235]

- 3 Xu K, Zhou P, Yang C L. On-line detection technique of tiny surface defects for metal plates and strips based on photometric stereo (in Chinese). *J Mech Eng*, 2013, 49: 25–29 [徐科, 周鹏, 杨朝霖. 基于光度立体学的金属板带表面微小缺陷在线检测方法. *机械工程学报*, 2013, 49: 25–29]
- 4 Feng Q C, Chen J, Li C L, et al. Measurement method of vegetable seedling leaf morphology based on photometric stereo (in Chinese). *Trans Chin Soc Agric Mach*, 2018, 49: 43–50 [冯青春, 陈建, 李翠玲, 等. 基于光度立体视觉的蔬菜秧苗叶片形态测量方法. *农业机械学报*, 2018, 49: 43–50]
- 5 Cai Y H, Zhang L L, Sheng N, et al. 3D reconstruction of tongue surface based on photometric stereo method (in Chinese). *J Electron Inform Tech*, 2015, 37: 2564–2570 [蔡轶珩, 张琳琳, 盛楠, 等. 基于光度立体法的中医舌体三维表面重建. *电子与信息学报*, 2015, 37: 2564–2570]
- 6 Hayakawa H. Photometric stereo under a light source with arbitrary motion. *J Opt Soc Am A*, 1994, 11: 3079–3089
- 7 Miyazaki D, Hara K, Ikeuchi K. Median photometric stereo as applied to the Segonko tumulus and museum objects. *Int J Comput Vis*, 2010, 86: 229–242
- 8 Wu L, Wang Y T, Liu Y. A robust approach based on photometric stereo for surface reconstruction (in Chinese). *Acta Autom Sin*, 2013, 39: 1339–1348 [吴仑, 王涌天, 刘越. 一种鲁棒的基于光度立体视觉的表面重建方法. *自动化学报*, 2013, 39: 1339–1348]
- 9 Georgiades A S. Recovering 3D shape and reflectance from a small number of photographs. In: *Proceedings of the 14th Eurographics Workshop on Rendering Techniques*. Leuven, 2003. 230–240
- 10 Goldman D B, Curless B, Hertzmann A, et al. Shape and spatially-varying BRDFs from photometric stereo. *IEEE Trans Pattern Anal Mach Intell*, 2010, 32: 1060–1071
- 11 Santo H, Samejima M, Sugano Y, et al. Deep photometric stereo network. In: *2017 IEEE International Conference on Computer Vision Workshop (ICCVW)*. Venice, 2017. 501–509
- 12 Chen G Y, Han K, Wong K-Y K. PS-FCN: A flexible learning framework for photometric stereo. In: Ferrari V, Hebert M, Sminchisescu C, et al., eds. *Computer Vision—ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, Vol 11213*. Cham: Springer, 2018. 3–19
- 13 Ikehata S. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In: Ferrari V, Hebert M, Sminchisescu C, et al., eds. *Computer Vision—ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, Vol 11219*. Cham: Springer, 2018. 3–19
- 14 Huang G, Liu Z, Weinberger K Q. Densely connected convolutional networks. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. 4700–4708
- 15 Matusik W, Pfister H, Brand M, et al. A data-driven reflectance model. *ACM Trans Graph*, 2003, 22: 759–769
- 16 Wang X, Ren M J, Xiao G B, et al. A data-driven photometric stereo algorithm (in Chinese). *Sci Sin Tech*, 2018, 48: 969–979 [王晰, 任明俊, 肖高博, 等. 一种基于高斯过程的数据驱动光度立体视觉全局优化算法. *中国科学: 技术科学*, 2018, 48: 969–979]
- 17 Johnson M K, Adelson E H. Shape estimation in natural illumination. In: *CVPR 2011*. Colorado Springs: IEEE, 2011. 2553–2560
- 18 Argyriou V, Petrou M. Chapter 1 Photometric stereo: An overview. *Adv Imag Electron Phys*, 2009, 156: 1–54
- 19 Gong C. Exploring commonality and individuality for multi-modal curriculum learning. In: *AAAI'17 Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. San Francisco: AAAI, 2017. 1926–1933
- 20 Gong C, Tao D, Maybank S J, et al. Multi-modal curriculum learning for semi-supervised image classification. *IEEE Trans Image Process*, 2016, 25: 3249–3260
- 21 Fu K, Gong C, Gu I Y H, et al. Normalized cut-based saliency detection by adaptive multi-level region merging. *IEEE Trans Image Process*, 2015, 24: 5671–5683
- 22 Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, 2015. 1–9
- 23 Wiles O, Zisserman A. SilNet: Single- and multi-view reconstruction by learning from silhouettes. arXiv: 1711.07888
- 24 Jakob W. Mitsuba renderer. 2010, <http://www.mitsuba-renderer.org/>
- 25 Ikehata S, Aizawa K. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, 2014. 2187–2194

Non-Lambertian photometric stereo vision algorithm based on a multi-scale convolution deep learning architecture

CHEN MingHan, REN MingJun, XIAO GaoBo, WANG Xi & FU Lin

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

As a form of fine 3D measurement technology, photometric stereo vision is widely used in 3D reconstruction, defect detection, biomedicine, and other fields, but the traditional reflection model has limited ability to reflect the real physical properties of materials and is not suitable for non-Lambertian surfaces with nonlinear light reflection characteristics such as bright metal, greatly limiting broader application of the technology. This paper proposes a photometric stereo vision algorithm based on a multi-scale convolution deep learning architecture that implements high precision reconstruction of a nontransparent material surface under arbitrary illumination conditions. The algorithm implements a multi-scale convolution structure in the deep network. The small-scale convolution kernel strengthens the expression of photometric physics that gives the model superior detail prediction performance. The large-scale convolution kernel encourages deep networks to take advantage of neighborhood features and enhances the ability of models to overcome shadows and distinguish multiple materials. To further address the problem of arbitrary illumination conditions, the algorithm employs a multi-resolution polar coordinate division method for the illumination vector space that integrates the input image information and fully utilizes the multi-scale convolution. It is experimentally demonstrated that the multi-scale convolution deep learning architecture effectively integrates the advantages of both photometric stereo and deep learning and greatly improves the adaptability to non-Lambertian surfaces while retaining the ability to perform fine reconstruction. It provides powerful support for a wide range of applications in photometric technology.

deep learning, photometric stereo, multi-scale convolution

doi: [10.1360/SST-2019-0164](https://doi.org/10.1360/SST-2019-0164)