

复杂环境下基于深度神经网络的摄像机标定

向鹏, 周宾*, 祝仰坤, 贺文凯, 岳晓庚, 陶依贝

东南大学能源与环境学院, 江苏 南京 210096

摘要 提出一种基于深度神经网络的摄像机标定方法, 实现了复杂环境下平面区域内的灵活、高精度标定。无需进行数据特征提取或分类, 仅通过优化网络结构、超参数与训练算法, 神经网络便能得到快速有效的训练。实验结果表明, 相较于张正友标定法与浅层神经网络标定法, 该方法在大范围、多拍摄角度和高畸变条件下均能达到更高的标定精度, 镜头存在高畸变时, 633 mm×763 mm 标定范围内的平均标定误差仅为 0.1471 mm。

关键词 机器视觉; 摄像机标定; 深度神经网络; 修正线性单元; 自适应矩估计

中图分类号 TP391; TP183

文献标识码 A

doi: 10.3788/LOP56.111505

Camera Calibration Based on Deep Neural Network in Complex Environments

Xiang Peng, Zhou Bin*, Zhu Yangkun, He Wenkai, Yue Xiaogeng, Tao Yibei

School of Energy and Environment, Southeast University, Nanjing, Jiangsu 210096, China

Abstract This study proposes a new deep neural network based camera calibration method that achieves flexible, high-precision calibration in complex environments, without having to classify or extract features from input data. By optimizing the network structure, hyperparameters, and training algorithms, the deep neural network can be quickly and effectively trained. The experimental results confirm that, compared with Zhang's calibration method and the shallow neural network, the proposed method can achieve high calibration accuracy under a wide range of imaging conditions involving multiple shooting angles or high distortion. For the images produced using a highly distorted lens, the proposed method achieves an average calibration error of only 0.1471 mm over the calibration range of 633 mm×763 mm.

Key words machine vision; camera calibration; deep neural network; rectified linear unit; adaptive moment estimation

OCIS codes 150.1488; 150.0155; 200.4260

1 引言

标定是将仪器或系统与测量标准进行比较进而确定其性能参数的过程。摄像机标定是计算机视觉中的重要步骤, 从二维图像信息中获取真实世界的度量信息^[1], 在三维重建、姿态估计和虚拟现实等领域具有广泛的应用。

摄像机通过透镜将三维场景投影到像平面上, 获得二维图像, 图像上每一点与空间物体表面相应点的位置关系可由摄像机模型算得。以针孔模型为

基础, 学者们对摄像机标定方法进行了大量研究^[2-6]。这些标定方法的主要原理为: 将成像过程视为线性变化和非线性变化的叠加, 即真实世界物体经镜头投影到摄像机成像平面以及实际成像与理想成像间存在光学畸变两个因素的叠加, 引入内参、外参和畸变系数等参数进行量化并求解。此类方法需要首先利用线性模型求取部分参数初值, 然后计算其他参数, 再对初值进行优化。然而, 当摄像机镜头畸变较大时, 初值求解会产生较大误差, 同时对镜头畸变项与噪声等因素考虑不完全, 易导致最终标定

收稿日期: 2018-12-20; **修回日期:** 2018-12-25; **录用日期:** 2019-01-07

基金项目: 国家重点研发计划(2017YFB0603204)、国家自然科学基金(50976024)、国家自然科学基金青年基金(50906013)

* **E-mail:** seuxp@foxmail.com

结果并不精确;且对于应用于自主导航、医学和监视等领域的摄像机(如反射折射摄像机或鱼眼摄像机),针孔模型已不再适用^[7]。

含有至少一个隐藏层的前馈神经网络具有通用逼近性质^[8],可用于摄像机标定,实现从真实世界到图像的非线性映射过程。隐藏层数较少的浅层神经网络(SNN)可应用于摄像机标定中。Wen等^[9]提出的混合标定方法将SNN用于针孔模型标定结果的优化,提高了标定精度,但SNN并未直接用于映射全过程;Jun等^[10]提出了一种训练数据集预分类法,并将SNN用于二维平面标定的全过程,取得了比两步法更精确的标定结果,但其仅对较小标定范围进行了验证;Memon等^[11]使用双目摄像机验证了SNN用于三维区域标定的可行性;吕朝辉等^[12]分析了SNN与摄像机标定结合的3种不同方法,并进行了实验验证;田震等^[13]提出采用特殊的归一化方法对数据集进行预处理,再使用双SNN并行方法进行标定,提高了标定精度,但手动进行特征提取和同时训练两个SNN使标定过程更复杂,降低了机器学习类方法进行标定的灵活性;刘小娟等^[14]采用单角点棋盘格构造的虚拟标定靶替代常规标定靶,并结合SNN实现了大视场区域的标定,但单角点逐个拍摄大幅增加了标定时间。

虽然SNN标定法能克服传统标定法中模型复杂的问题,在较简单标定环境下标定精度有所提升,但以上研究大多未验证所采用方法的应用范围,标定结果多为简单环境,如小范围、摄像机像平面与标定平面接近平行、摄像机镜头无畸变或畸变较小等。当标定范围较大,拍摄角度变化或镜头存在高畸变时,反映成像过程的非线性映射函数更复杂,而SNN训练过程并不稳定,非线性映射能力较低,且需要对输入数据进行归一化,进而使标定流程更复杂。深度学习是机器学习中一种对大量数据进行表示学习的方法,允许由多个处理层组成的计算模型学习具有多个抽象级别的数据表示^[15]。深度神经网络(DNN)是隐藏层数大于1的前馈型神经网络^[16],其每一层能对前一层的抽象表示进行深入学习,无需手工进行数据特征提取,能比SNN学习到更复杂的函数,为深度学习网络模型之一。近年来,随着计算硬件、网络结构与训练算法的优化,已经能实现对DNN的有效训练,但目前将深度学习技术应用于摄像机标定的研究较少。

受SNN标定法与深度学习技术启发,本文提出一种基于DNN的摄像机标定方法。该方法结合

物理成像过程特点和DNN更强的通用逼近性质,构建适用于摄像机标定的DNN后,对训练算法进行优化,使其能够有效训练,无需手动进行数据集特征提取,直接输入原始图像数据,可以实现从真实世界到图像的非线性映射,在大范围、多拍摄角度和高畸变情况下均具有较高的标定精度,可用于不同的相机模型。

本研究的意义有:1)将深度学习技术应用于摄像机标定,对适用于摄像机标定的DNN结构、超参数选取与优化训练过程进行了详细分析与说明,使其能够得到快速有效的训练;2)在网络输入和输出层引入齐次项,使网络模型更接近射影几何物理模型,直接输入原始图像数据而不对其进行任何手工特征提取或分类,简化了标定流程;3)针对SNN标定法仅用于简单标定环境的问题,将机器学习类方法应用于大范围、多拍摄角度和高畸变等复杂标定环境中,并验证其可行性,结果表明,标定精度提高。

2 理论基础

2.1 针孔相机模型

空间中任意一点 P ,经透视投影后成像于像平面上的点 P' ,可用图像像素坐标系、图像物理坐标系、摄像机坐标系及世界坐标系4个参考坐标系定量描述两点之间的位置关系。假定 P 点位于世界坐标系 $z_w=0$ 平面上,其在图像像素坐标系、图像物理坐标系和世界坐标系中的坐标分别为 (u, v) 、 (x, y) 和 $(x_w, y_w, 0)$,则3者之间的关系可由齐次坐标表示为

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (1)$$

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{21} & t_1 \\ r_{12} & r_{22} & t_2 \\ r_{13} & r_{23} & t_3 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}, \quad (2)$$

式中: dx 、 dy 分别为单位像素在 x 、 y 轴上的物理尺寸; (u_0, v_0) 为图像物理坐标系的原点在图像像素坐标系下的坐标; s 为比例因子; f 为摄像机焦距; \mathbf{r}_1 (r_{11}, r_{12}, r_{13})、 \mathbf{r}_2 (r_{21}, r_{22}, r_{23})为旋转向量; \mathbf{t} (t_1, t_2, t_3)为平移向量。

然而,实际的镜头并非理想的透视成像,而是带有不同程度的畸变,因此,实际成像位置不是针孔模型所描述的图像物理坐标位置 (x, y) ,而是受到镜头失真影响发生偏移后的位置 (x', y') ,其关系为

$$\begin{cases} x = x' + \delta_x(x', y') \\ y = y' + \delta_y(x', y') \end{cases} \quad (3)$$

式中： δ_x 与 δ_y 为非线性畸变项。

联立(1)式、(2)式和(3)式,可得畸变情况下摄像机实际拍得图像上的点(u', v')与世界坐标系点(x_w, y_w)的关系为

$$\begin{cases} (u' - u_0)dx + \delta_x(x', y') = \frac{f(r_{11}x_w + r_{21}y_w + t_1)}{s(r_{13}x_w + r_{23}y_w + t_3)} \\ (v' - v_0)dy + \delta_y(x', y') = \frac{f(r_{12}x_w + r_{22}y_w + t_2)}{s(r_{13}x_w + r_{23}y_w + t_3)} \end{cases} \quad (4)$$

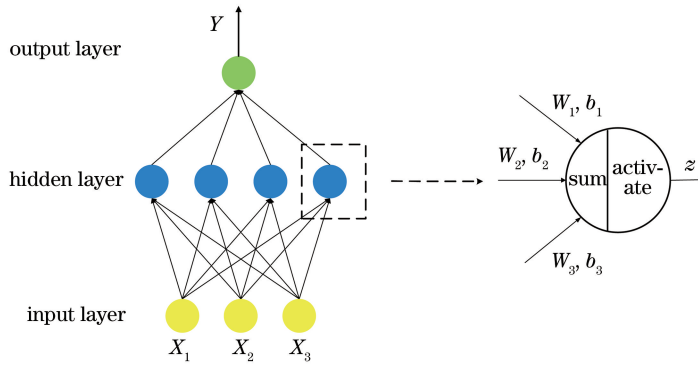


图 1 浅层神经网络模型
Fig. 1 Model of shallow neural network

为使 SNN 的输出尽可能逼近目标值,需要采用有监督学习的方法。输入网络带有标签的数据,损失函数 L 衡量网络输出值与标签值之间的差异,其值越高表示差异越大,并使用梯度下降(GD)法对每个神经元的权重与偏置进行反复迭代更新,使其收敛至预定的目标范围。

2.3 DNN 模型

DNN 隐藏层数较 SNN 更多,拥有比 SNN 更复杂的网络结构。随着隐藏层数增加,网络容量增加,神经元之间更多的协作可以学习到更复杂的函数,也能容纳更多的数据量。但网络层数的增加会导致梯度反向传播过程易发生梯度弥散或梯度爆炸,从而使网络难以得到有效训练,同时,层数增加导致参数急剧增多,计算成本也极大增加。目前,训练 DNN 的方法主要有:

1) 无监督逐层贪婪预训练方法^[17]。逐层无监督预训练为 DNN 提供较好的初始参数,每次训练一层隐藏节点,训练时上一层隐节点的输出为输入,本次隐节点的输出为下一层隐节点的输入;完成各层预训练后,结合 GD 的反向传播算法对整个网络进行微调。

2.2 SNN 模型

由于计算机硬件和算法的原因,早期只有隐藏层数较少的 SNN 能够有效训练。典型的 SNN 模型如图 1 所示,其由输入层、隐藏层和输出层组成,其中 W, b 和 z 分别为神经元的权重、偏置和输出。

SNN 模型中输出 Y 与输入向量 X 的关系为

$$z = a(W_{input}X + b_{input}), \quad (5)$$

$$Y = W_{hidden}z + b_{hidden}, \quad (6)$$

式中: z 为隐藏层输出向量; W_{input}, b_{input} 分别为输入层权重矩阵与偏置向量; a 为 sigmoid 或 tanh 激活函数; W_{hidden}, b_{hidden} 分别为隐藏层权重矩阵与偏置向量。

2) 改进的训练算法。对网络结构及训练算法进行改进与优化,并使用图形处理单元(GPU)对训练过程进行加速。

相较于随机或利用各种算法初始化权值矩阵,方法 1 进行分层预训练具有一定的优势,但每层贪婪学习权值矩阵易导致训练时间过长。随着 GPU 等硬件的使用和网络结构、算法的优化,改进的训练算法的训练效果已经优于预训练方法^[18]。本文使用方法 2 对 DNN 进行训练,具体过程如下:

1) 定义网络框架,随机初始化网络所有神经元的参数值,定义损失函数 L 。

2) 输入训练数据集,使用前向传播算法计算网络 L 值。

3) 使用改进算法逐层更新神经元权重和偏置值。

4) 重复步骤 2)~3),直至输出 L 值收敛至预定范围。

对于(7)式所示的非线性转换,可利用 DNN 强大的通用逼近能力,直接实现从世界坐标系到图像像素坐标系的非线性映射。

3 超参数选取与算法改进

3.1 DNN 框架

3.1.1 数据集

当给予学习模型更多的数据时,能获得更好的结果,DNN 拥有比 SNN 更强的大数据处理能力,在描述大数据内在逻辑方面具有优势。选取平面内 90×110 个共 9900 个点作为数据集,其中随机选取 9000 个点为训练数据集,剩余 900 个点为测试数据集。

3.1.2 网络框架

所用 DNN 框架包括输入层、输出层和 4 个隐

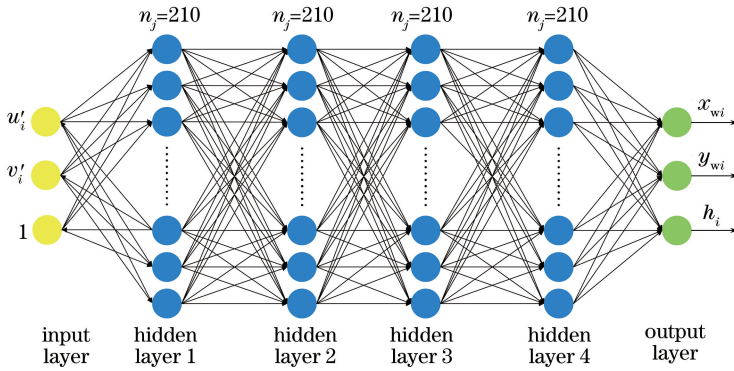


图 2 DNN 框架

Fig. 2 Framework of DNN

3.1.3 参数初始化

参数的初始化值会严重影响 DNN 的学习过程。由链式法则可知,权重初始化值过小或过大均会导致梯度弥散或梯度爆炸,从而使 DNN 无法收敛或收敛速度过慢;因网络为全连接结构,每层单个神经元受上一层所有神经元的影响,若参数初始化值都相同,将会使每个神经元输出相同值,无法打破对称性。为克服以上问题,随机初始化每层神经元权重值并使其满足标准高斯分布,由于输入为图像像素坐标值,数量级为 10^3 ,因此对权重值进行一定比例的缩小(缩小因子 $\eta = 5/n_j$);随机初始化权重值已打破对称性,初始化所有神经元偏置值为 0。

3.1.4 损失函数

假设网络第 i 个训练样本点的输出值为 (x'_{wi}, y'_{wi}, h_i) ,以均方误差定义损失函数,则

$$L = \frac{1}{M} \sum_{i=1}^M (x'_{wi} - x_{wi})^2 + (y'_{wi} - y_{wi})^2 + (h_i - 1)^2, \quad (7)$$

式中: M 为训练样本点总个数; $(x_{wi}, y_{wi}, 1)$ 为第 i 个训练样本点在世界坐标系下的真实齐次坐标。

藏层,如图 2 所示。其中输入输出层与一般用于摄像机标定 SNN 的输入输出层不同。为使框架更接近摄像机物理成像过程,依据射影几何原理,将输入输出坐标增加一个维度后转换为齐次坐标,故输入输出层分别为图像像素坐标系与世界坐标系对应点的齐次坐标 $(u', v', 1)$ 和 (x'_w, y'_w, h) ,其神经元数均为 3。直接将原始图像像素坐标数据输入网络,不对其进行任何特征提取或分类。隐藏层每层的神经元个数 n_j 均为 210,故所构建的 DNN 框架结构为 $3 \times 210 \times 210 \times 210 \times 210 \times 3$ 。

3.1.5 激活函数

由于 DNN 隐藏层数较多,利用标准反向传播算法对每层权重、偏置进行更新时,使用 sigmoid 或 tanh 函数存在以下缺点:sigmoid 或 tanh 函数的导数总是小于 1,且随输入值的绝对值增大而减小,经多层反向传播后,梯度更新信息接近于 0,进而导致梯度弥散,参数停止更新;sigmoid 或 tanh 函数正向和反向传播过程中均涉及指数运算,当网络层数与神经元个数较多时,计算成本大幅增加。而修正线性单元(ReLU) $f(x) = \max(0, x)$ 具有分段线性,在稀疏输出和反向传播过程中不易造成梯度更新信息丢失,特别适合本文所使用的 DNN 框架,故将其作为每层神经元的激活函数。

3.2 参数更新算法优化

如图 3 所示,在使用 GD 法对网络权重、偏置进行更新时,只有当学习速率 $\alpha \leq 1 \times 10^{-6}$ 时, L 才能逐渐收敛,且始终无法收敛至理想范围。这是因为网络框架越大,鞍点越多,GD 法对参数更新时只考虑当次迭代的更新信息,当迭代过程到达鞍点时,当次迭代的梯度接近于 0,参数始终无法更新,训练陷入停滞区域。

为了使训练过程能跳出鞍点并尽快收敛至理想范围,本文采用自适应矩估计(Adam)法对参数进行更新,某一神经元的权重在第 t 次迭代后的值 w_t 的更新过程为

$$m_t = [\beta_1 m_{t-1} + (1 - \beta_1)(\partial L / \partial w_{t-1})] / (1 - \beta_1'), \quad (8)$$

$$v_t = [\beta_2 v_{t-1} + (1 - \beta_2)(\partial^2 L / \partial w_{t-1}^2)] / (1 - \beta_2'), \quad (9)$$

$$w_t = w_{t-1} - \alpha m_t / (\sqrt{v_t} + \epsilon), \quad (10)$$

式中: β_1 、 β_2 为指数衰减率,本文取 $\beta_1 = 0.9$, $\beta_2 = 0.99$; t 为迭代次数; m_t 为第 t 次梯度指数移动均值, $m_0 = 0$; v_t 为第 t 次平方梯度, $t_0 = 0$; α 为学习速率; ϵ 为常数,本文取 $\epsilon = 1 \times 10^{-8}$ 。

由(8)~(10)式可知,Adam 法在每次迭代过程中充分考虑前次迭代梯度信息,使更新参数拥有动量因素 m_t ,引入平方梯度项 v_t 以减小收敛过程中的摆动,故学习速率 α 可以更大,进而使训练过程可以尽快跳出鞍点并接近全局最优点,其效果如图 3 所示。

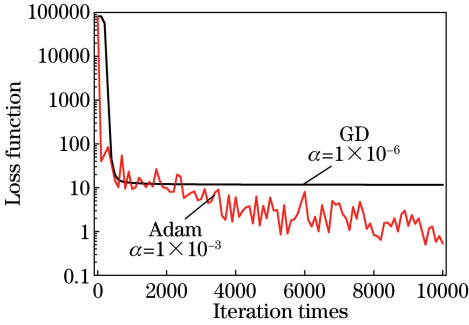


图 3 不同参数更新算法下的损失函数收敛曲线

Fig. 3 Loss function convergence curves under different parameter update algorithms

3.3 超参数选取

一些超参数值的选取对 L 收敛速度和精度的影响很大,甚至可以直接导致其无法收敛。现对这些超参数进行分析,为减少权重随机初始化的影响,结果均为重复 10 次后的平均值。

3.3.1 学习速率

参数每次迭代更新过程均受到学习速率 α 的影响。 α 太小会导致收敛速度很慢,以致长时间无法收敛至理想精度; α 太大会导致 L 无法收敛甚至扩散。本文取 $\alpha = [10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}]$ 对网络进行训练, L 随迭代次数的变化如图 4 所示。

由图 4 可知,当 $\alpha \geq 1 \times 10^{-1}$ 时, L 无法收敛;随着 α 的减小, L 开始收敛,当 $\alpha = 1 \times 10^{-3}$ 时,达到最

快收敛速度,但收敛过程剧烈抖动;当 $\alpha \leq 1 \times 10^{-4}$ 时,收敛速度逐渐放缓,但收敛过程趋于稳定。为使训练前期具有较大的收敛速度,后期能够精确收敛至理想范围,取学习速率初始值为 $\alpha_0 = 6 \times 10^{-3}$,并使其随迭代次数的增加呈阶梯式指数衰减,即

$$\alpha_t = 0.7^{\text{floor}(t/1000)} \alpha_0, \quad (11)$$

式中: α_t 为第 t 次迭代时的学习速率; $\text{floor}(\cdot)$ 为向下取整函数。

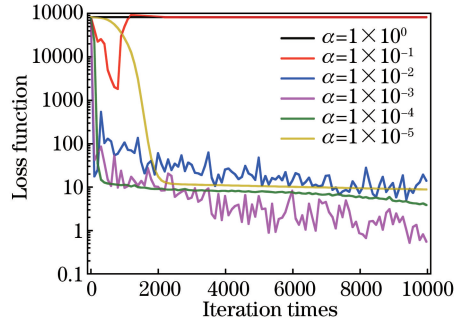


图 4 不同学习速率下损失函数收敛曲线

Fig. 4 Loss function convergence curves under different learning rates

3.3.2 隐藏层数及神经元数

隐藏层越多,网络具有越深入的特征表示能力,同时神经元数的增多也会增加网络总参数量,整体上均能够增强网络的容量,使其能学习到更复杂的函数。但隐藏层数与神经元数的增多会导致计算成本增加,甚至出现过拟合情况。将这两个超参数置于二维网格内,对最优取值进行试探性搜索,随机搜索法具有比网格搜索法更高的效率^[19],故使用随机搜索法确定超参数值。如图 5 所示,假定每层神经元数相等,随机选取隐藏层数为 2~10,每层神经元数为 10~210(取点间隔为 20),共选取 25 个试验点。

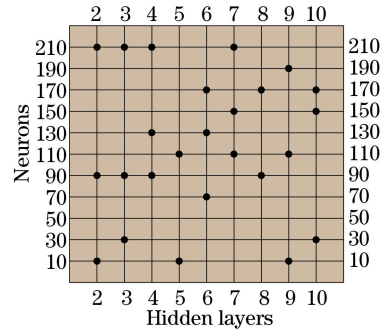


图 5 随机搜索的 25 个试验点

Fig. 5 Twenty-five trials for random search

使用每个试验点所代表的隐藏层和神经元数值构建 DNN,其他参数及结构如前文所示。经 15000

次迭代后,25个试验点的 L 值及计算耗时如图6所示。由图6可知,隐藏层和神经元数较少时,无法使 L 收敛至较小的范围,当隐藏层数为4,每层神经元数为210时, L 取得最小值,此后,随着隐藏层数的增加, L 不减反增,且训练时间急剧增加。因此,综合考虑 L 收敛程度和训练时间后,取隐藏层数为4,每层神经元个数为210。

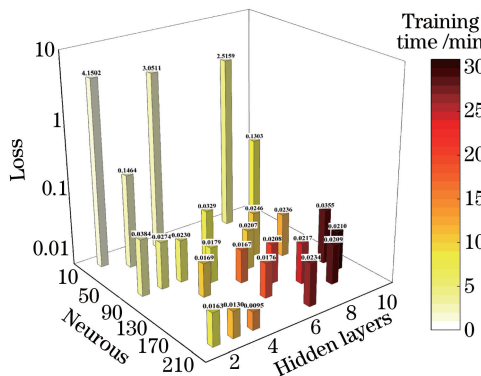


图6 试验点迭代后的损失及训练用时

Fig. 6 Loss and training time of test points after iteration

4 实验与结果分析

实验所用标定靶为 90×110 平面角点阵列,如图7所示,其中相邻角点间距均为7 mm,故标定区域范围为 $623 \text{ mm} \times 763 \text{ mm}$ 。实验过程中,随机选择范围内9000个点为训练数据集,剩余900个点为测试数据集。

固定摄像机及标定靶位置,采集单张图像,利用Matlab角点检测函数得到所有角点的图像像素坐标(以图像左上角为原点)。为满足输入数据的正负均匀性,将坐标系原点平移至角点阵列中心位置,得到所有角点图像像素坐标与真实世界坐标的一一对应关系,将坐标转换为齐次坐标后即可得到数据集及其标签。

使用Google公司深度学习开源软件库Tensorflow构建和训练DNN。Tensorflow采用数据流图进行数值计算,具有高度的灵活性与可移植性,支持多种编程语言,能够自动计算微分,可以充分应用计算设备的GPU等硬件的性能,相较于传统方式,Tensorflow能使DNN的构建过程更简洁,训练速度更快。一旦DNN训练完毕,将待测平面内任意点图像像素坐标经齐次转换后输入DNN,即可得到真实世界下的位置坐标。

为验证DNN标定法在不同应用场景下的性

能,在相同实验条件下,比较张正友标定法、SNN标定法和DNN标定法的标定结果,其中SNN为刘小娟^[11]所用的 $2 \times 13 \times 2$ 网络结构,激活函数为tanh,为适应激活函数性质,对所有输入数据及其对应标签数据进行归一化预处理。实验结果均为10次随机选取测试数据集点标定结果的平均值。以标定误差 e_i 作为评价指标,其表达式为

$$e_i = \sqrt{(x'_{wi} - x_{wi})^2 + (y'_{wi} - y_{wi})^2}, \quad (12)$$

式中: (x'_{wi}, y'_{wi}) 和 (x_{wi}, y_{wi}) 分别为第 i 个测试数据集点在世界坐标系下的测量坐标与真实坐标。

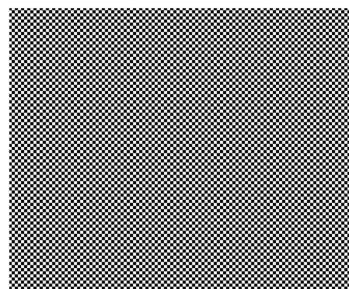


图7 标定靶

Fig. 7 Calibration target

4.1 拍摄角度

受一些特殊的应用场景的限制,摄像机成像平面与待测平面间的角度只能固定在一定范围内。假设两平面平行时, $\theta = 0^\circ$,将摄像机沿摄像机坐标系 y_c 轴逐次旋转 15° ,每次旋转后对待测平面进行标定。表1给出了不同标定方法在不同拍摄角度下,900个测试数据集点的平均标定误差(Avg)、最大标定误差(Max)和误差标准差(Std)。

由表1结果可知,当标定范围较大时,SNN标定法标定误差较大,这是因为当标定范围增大时,为了将更精确的先验信息输入至网络,需要增加数据集点数,但SNN模型复杂度较低,较大的数据量易使训练过程陷入局部最优;张氏标定法标定误差与拍摄角度无关,其标定误差取决于针孔模型中各个参数的标定精度,拍摄角度的变化只对外参中旋转矩阵的数值产生影响,但内误差标准差相对较大,可能是仅考虑径向畸变等原因,对物理成像过程描述不完全,当对大范围内多个点进行标定时,随机选取的不同验证点位于不同的成像区域导致误差波动较大;DNN标定法于所有拍摄角度下的标定误差均明显小于其他方法,具有较高的标定精度,且误差波动较小,证明本文设计的深度神经网络框架从训练数据集中有效学习到了从真实世界到图像的非线性映射。此外,可以发现,DNN标定法和SNN标定法

的平均标定误差随拍摄角度的增加而增加,证明非线性映射函数复杂度随拍摄角度变化而变化,SNN 标定法标定误差的变化程度更大,当拍摄角度为

60°时,其最大标定误差达到 20.3086 mm,已无法对成像过程进行有效学习;DNN 标定法标定误差的变化程度较小,稳定性较强。

表 1 不同方法在不同拍摄角度下的标定误差

Table 1 Calibration errors of different methods under different shooting angles

Angle / (°)	Error of DNN method / mm			Error of SNN method / mm			Error of Zhang's method / mm		
	Avg	Max	Std	Avg	Max	Std	Avg	Max	Std
0	0.0896	0.4109	0.0553	1.8910	8.4989	1.3840	0.3887	2.2614	0.2714
15	0.0948	0.5774	0.0657	3.0027	9.1213	1.6706	0.4768	1.7119	0.3339
30	0.1271	0.7146	0.0895	3.3793	11.8804	2.1458	0.3722	3.0325	0.2733
45	0.1492	0.7831	0.0930	3.9889	15.5627	2.6222	0.2586	1.7789	0.2139
60	0.1897	0.8929	0.1003	4.6515	17.3086	2.8226	0.5350	3.2883	0.4284

4.2 镜头高畸变

摄像机镜头畸变对标定结果具有很大影响,对于某些特殊应用的摄像机甚至针孔模型也无法对其成像过程进行准确描述。鱼镜头头具有很大的视角范围,广泛应用于场景监视、卫星定位、机器人导航、微小智能系统及工程测量等领域,但视角的增大会引入较大的畸变,主要表现为桶形畸变。本文使用装有鱼镜头头的摄像机,在 $\theta=15^\circ$ 条件下对待测平面进行标定,如图 8 所示。表 2 给出了不同标定方法的标定误差,其中张氏标定法的标定误差为进行畸变校正后的结果。

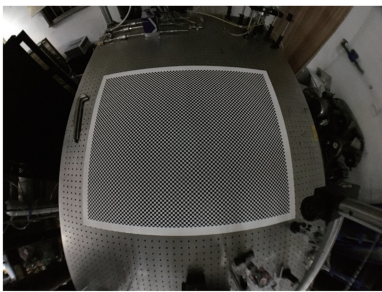


图 8 高畸变图像

Fig. 8 Highly distorted image

表 2 不同方法在镜头高畸变下的标定误差

Table 2 Calibration errors of different methods with highly distorted lens mm

Error	DNN method	SNN method	Zhang's method
Avg	0.1471	7.8896	2.8714
Max	0.8990	34.2669	7.1937
Std	0.1124	4.8628	1.7055

由表 2 结果可知,镜头存在高畸变时,反映成像过程的非线性映射函数更复杂,SNN 的学习能力有限,已无法达到理想的标定精度;即使进行了畸变校正,因针孔模型对物理成像过程的描述不完全,张氏标定法标定误差依旧较大;DNN 标定法几乎不受镜

头畸变的影响,标定误差明显小于其他方法,平均标定误差仅为 0.1471 mm,说明 DNN 标定法适用于不同的相机模型,再次验证了 DNN 对复杂非线性函数的强学习能力。激活函数对每一个神经元的作用过程代表一次非线性变换,DNN 隐藏层数的增加不仅增加了拥有激活函数的神经元数目,而且还增加了激活函数嵌套的层数,因此,DNN 拥有更强的通用逼近性质,能学习到更复杂的函数。

比较表 1 和表 2 中 DNN 标定法于 $\theta=15^\circ$ 条件下的标定结果可知,镜头高畸变增加了非线性映射函数的复杂度;镜头存在高畸变时,DNN 标定法误差波动相对较大,这可能是学习更复杂的函数时,网络出现过拟合所致。

4.3 齐次坐标

在 $\theta=15^\circ$,无畸变的条件下,比较加入齐次项与不含齐次项的 DNN 标定效果。不同结构的 DNN 标定误差如表 3 所示。

表 3 不同结构的 DNN 标定误差

Table 3 Calibration errors of DNN with different structures mm

Error	With	Without
	homogeneous term	homogeneous term
Avg	0.0948	0.1325
Max	0.5774	0.8654
Std	0.0657	0.0992

由表 3 的结果可知,网络输入和输出层引入齐次项后,标定精度能够进一步的提升。对于物理模型,加入齐次项使输入输出过程更接近射影几何成像原理;而对于 DNN,齐次项的引入增加了一个维度的输入信息,故能使网络更好地从数据中学习特征。

需要强调的是,所使用的深度神经网络框架相较于 SNN 模型,其之所以能得到较高的精度,本质

上不仅是因为网络层数和神经元数的增加,而且还有激活函数的改进、参数更新算法的优化以及 GPU 的使用。这些原因可使更复杂的网络得到快速有效的训练,进而收敛至理想范围内,这也是深度学习技术和传统神经网络的区别之一。若保持隐藏层数与神经元个数不变,忽略输入输出层齐次项,使用传统神经网络 sigmoid 或 tanh 激活函数,并用 GD 法进行参数迭代更新,因存在数据预处理步骤,不仅会使标定流程复杂化,而且还会使网络训练时间与标定误差急剧增大。经实验验证,在 $\theta = 15^\circ$, 无畸变条件下,上述结构神经网络的平均标定误差为 8.8093 mm,训练时间为 40.8 min,效果甚至不如 SNN 结构。

5 结 论

提出了一种基于 DNN 的大范围摄像机标定方法,通过构建特定的 DNN 框架,确定超参数并对训练算法进行优化,使网络得到有效训练,实现大范围平面区域的高精度标定。该方法利用 DNN 更强的通用逼近能力,直接输入原始图像坐标数据而无需进行任何特征提取或分类,使用 GPU 实现对网络训练过程的加速,实现从真实世界到图像的非线性映射过程,适用于不同的相机模型。实验结果表明,在大范围、不同拍摄角度和镜头存在高畸变的复杂标定环境下,所提方法能取得比张友标定法与 SNN 标定法更高的标定精度,具有较高的灵活性与稳定性,特别适用于无需摄像机内外参信息,仅需通过图像获取真实世界大范围度量信息的应用场景。同时,实验结果对双目视觉立体标定具有一定的理论借鉴意义。

参 考 文 献

- [1] Zhang Z Y. Camera calibration [M] // Ikeuchi K, Computer Vision: A Reference Guide. Boston: Springer, 2014: 76-77
- [2] Tsai R. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses[J]. IEEE Journal on Robotics and Automation, 1987, 3(4): 323-344.
- [3] Zhang Z. A flexible new technique for camera calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
- [4] Zou P P, Zhang Z L, Wang P, *et al.* Binocular camera calibration based on collinear vector and plane homography[J]. Acta Optica Sinica, 2017, 37(11): 1115006.
- [5] Wu Q H, Chen H, Zhu S S, *et al.* Fast calibration method of large dimension measurement system based on multi-camera array[J]. Acta Optica Sinica, 2018, 38(12): 1215002.
- [6] Zhu F, Yu F S, Wu Y M, *et al.* Analysis of attitude calibration precision of P4P camera[J]. Acta Optica Sinica, 2018, 38(11): 1115005.
- [7] Ramalingam S, Sturm P. A unifying model for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(7): 1309-1319.
- [8] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators [J]. Neural Networks, 1989, 2(5): 359-366.
- [9] Wen J, Schweitzer G. Hybrid calibration of CCD cameras using artificial neural nets[C] // 1991 IEEE International Joint Conference on Neural Networks, November 18-21, 1991, Singapore. New York: IEEE, 1991: 337-342.
- [10] Jun J, Kim C. Robust camera calibration using neural network[C] // Proceedings of IEEE. IEEE Region 10 Conference. TENCON 99. 'Multimedia Technology for Asia-Pacific Information Infrastructure' (Cat. No. 99CH37030), September 15-17, 1999, Cheju Island, Korea. New York: IEEE, 1999: 694-697.
- [11] Memon Q, Khan S. Camera calibration and three-dimensional world reconstruction of stereo-vision using neural networks [J]. International Journal of Systems Science, 2001, 32(9): 1155-1159.
- [12] Lü C H, Zhang Z Y, An P. Camera calibration for stereo vision based on neural network [J]. Chinese Journal of Mechanical Engineering, 2003, 39(9): 93-96.
- [13] Tian Z, Zhang Q, Xiong J L, *et al.* Large-scale camera calibration with neural network [J]. Acta Optica Sinica, 2011, 31(4): 0415001.

- 田震, 张玘, 熊九龙, 等. 基于神经网络的大范围空间标定技术[J]. 光学学报, 2011, 31(4): 0415001.
- [14] Liu X J, Li X J, Wang W Y, *et al.* Large-scale binocular camera calibration combining neural network with virtual target[J]. Optical Technique, 2017, 43(4): 314-318.
- 刘小娟, 李学军, 王文韞, 等. 基于神经网络的虚拟靶标大视场双目相机标定技术[J]. 光学技术, 2017, 43(4): 314-318.
- [15] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [16] Hinton G, Deng L, Yu D, *et al.* Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [17] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.
- [18] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]// Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, April 11-13, 2011, Fort Lauderdale, FL, USA. Cambridge: PMLR, 2011, 15: 315-323.
- [19] Bergstra J, Bengio Y. Random search for hyperparameter optimization [J]. Journal of Machine Learning Research, 2012, 13(1): 281-305.