

# 基于神经网络的大范围空间标定技术

田 震 张 玘 熊 九 龙 王 国 超

(国防科学技术大学机电工程与自动化学院, 湖南 长沙 410073)

**摘要** 目前利用神经网络进行摄像机标定的研究, 只能实现单个位置上或小范围空间的标定精度要求。当标定空间扩大到一定范围, 标定精度和标定速度的冲突便不可调节。首先推导了摄像机的成像原理, 证明沿世界坐标系 3 个坐标轴( $X_w$  轴,  $Y_w$  轴和  $Z_w$  轴)方向存在截然不同的成像规律, 提出沿上述 3 个方向分别进行标定的并行标定方法; 进而提出一种新的归一化方法, 较好的提高了  $X_w$  轴和  $Y_w$  轴方向的标定精度。实验结果表明,  $Z_w$  轴向标定是整个大范围标定的关键, 其重构标准差远大于  $X_w$  轴和  $Y_w$  轴的重构标准差。在保证精度与速度的前提下, 新的归一化方法扩大了摄像机标定范围。

**关键词** 机器视觉; 摄像机标定; 并行标定; 神经网络

**中图分类号** TP391 **文献标识码** A **doi:** 10.3788/AOS201131.0415001

## Large-Scale Camera Calibration with Neural Network

Tian Zhen Zhang Qi Xiong Jiulong Wang Guochao

(Mechatronics and Automation School, National University of Defense Technology, Changsha, Hunan 410073, China)

**Abstract** Current researches on camera calibration using neural network can only keep precision within a small space, unfortunately, there has no such a method that can solve conflict between calibration precision and speed as space extends. It is proved that there are different imaging rules along three axes ( $X_w$  axis,  $Y_w$  axis and  $Z_w$  axis) in the world coordinate system, and a brand new parallel calibration method is created to calibrate camera along these three directions respectively, namely parallel calibration. A new normalization method is proposed and better calibration precisions are gained along both  $X_w$  axis and  $Y_w$  axis. There comes conclusion that calibration along  $Z_w$  axis is critical to the whole large-scale calibration since standard deviation of reconstruction along  $Z_w$  axis is much higher than those along  $X_w$  axis and  $Y_w$  axis. Experimental results show that the new normalization method extends calibration scale while keeping both calibration precision and speed.

**Key words** machine vision; camera calibration; parallel calibration; neural network

**OCIS codes** 150.1488; 150.0150; 100.4996; 200.4260

## 1 引 言

计算机视觉的基本任务是从摄像机获取的图像信息中计算恢复出物体的三维空间信息。三维物点与其对应的二维像点之间的对应关系是由摄像机成像几何模型决定的, 确定这种对应关系的过程, 即摄像机标定。目前的标定方法可分为传统标定方法<sup>[1~4]</sup>、自标定方法<sup>[5]</sup>和基于主动视觉的标定方法<sup>[6]</sup>。其中基于神经网络的摄像机标定属于传统标定方法。

20 世纪 80 年代初, J. J. Hopfield<sup>[7]</sup> 和 D.

Rumelhart 等<sup>[8]</sup> 研究发现神经网络潜力巨大, 促使该领域的研究快速发展。鉴于神经网络在函数逼近方面的突出能力, J. Wen 等<sup>[9]</sup> 和 M. Lynch 等<sup>[10]</sup> 于 20 世纪 90 年代初开始将神经网络应用到摄像机标定技术当中。将神经网络应用到摄像机标定中, 可以跳过求取摄像机内外参数的过程, 利用像点和相应物点作为输入输出训练集进行训练, 得到能够反映输入输出映射关系的网络。

成像过程中存在线性因素和非线性因素: 光沿直线传播, 形成针孔成像模型, 此为线性因素; 而由

**收稿日期:** 2010-09-16; **收到修改稿日期:** 2010-10-25

**作者简介:** 田 震(1984—), 男, 硕士研究生, 主要从事摄像机标定方面的研究。E-mail: tzh552911@126.com

**导师简介:** 张 玘(1959—), 男, 教授, 博士生导师, 主要从事仪器科学与技术等方面的研究。

E-mail: zhang\_qi\_7233@163.com

于制造工艺及其他因素的影响,成像过程中存在径向及切向畸变等,这些是非线性因素。线性因素与非线性因素,前者为主导。后者决定径向识别精度<sup>[11,12]</sup>,目前的标定方法,大都依赖神经网络的非线性泛化能力,跳过了线性因素的研究,将研究重点直接放在非线性因素的研究与补偿上,从而导致其只能局限于单个位置或小范围空间的标定。随着标定空间范围的扩大,线性因素的主导地位体现出来后,其精度就无法继续保持下去。

本文从摄像机成像的基本原理出发,指出沿  $X_w, Y_w, Z_w$  三轴向具有不同的成像规律;提出了一种新的并行标定方法,建立不同的训练点集对世界坐标中的  $X_w, Y_w, Z_w$  信息分别进行标定,在保证总体标定精度与标定速度的同时扩大了标定空间;同时提出了一种新型归一化方法,进一步提高了  $X_w$  轴和  $Y_w$  轴的标定精度。

## 2 基于神经网络的大范围空间标定

### 2.1 线性成像规律分析

摄像机标定的最终任务,是建立起像点坐标  $(u, v)$  与物点世界坐标  $(X_w, Y_w, Z_w)$  之间的映射关系<sup>[13]</sup>。像点坐标  $(u, v)$  表示该像点在图像数组中的行数与列数,并没有用物理单位表示出该像点在图像中的位置,需要再建立以物理单位表示的图像坐标系  $XO_1Y$ 。其中  $O_1$  定义为摄像机光轴与图像平面的交点,如图 1 所示。齐次坐标  $(u, v, 1)^T$  与  $(X, Y, 1)^T$  之间有如下对应关系:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dX} & 0 & u_0 \\ 0 & \frac{1}{dY} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \quad (1)$$

式中  $(u_0, v_0)$  为  $O_1$  在  $uO_0v$  坐标系中的坐标,  $dX, dY$  分别为每个像素在  $X$  轴与  $Y$  轴方向上的物理尺寸。

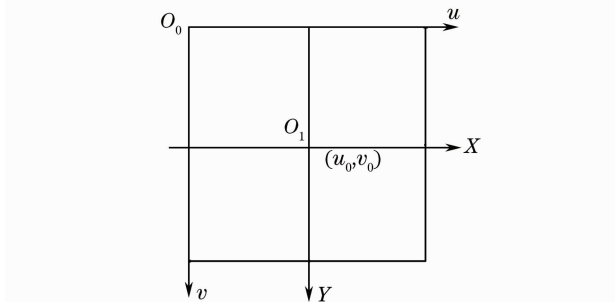


图 1 图像坐标系

Fig. 1 Image coordinate system

摄像机成像几何关系如图 2 所示。其中  $O$  点为摄像机光心,  $x$  轴和  $y$  轴与图像的  $X$  轴和  $Y$  轴平行,  $z$  轴为摄像机的光轴。点  $O$  与  $x, y, z$  轴组成摄像机坐标系,  $OO_1$  为摄像机焦距。世界坐标系由  $X_w, Y_w, Z_w$  轴组成,用来描述摄像机在空间的具体位置。设空间中某一点  $P$  在世界坐标系与摄像机坐标系下的齐次坐标分别为  $(X_w, Y_w, Z_w, 1)^T$  和  $(x, y, z, 1)^T$ , 二者存在如下关系:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (2)$$

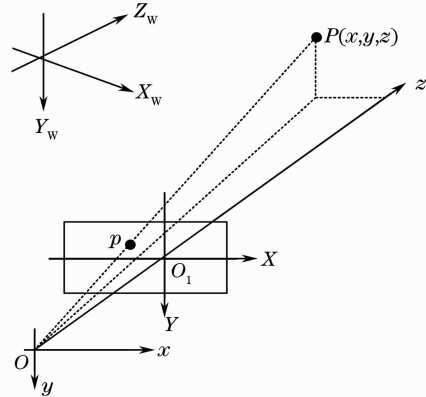


图 2 针孔成像模型

Fig. 2 Pinhole imaging model

如图 2 所示,物点  $P$  在摄像机坐标系下坐标  $(x, y, z)$  与对应图像坐标  $(X, Y)$  直接反映的是摄像机的成像几何关系。不考虑非线性因素时,可以通过针孔成像模型来描述成像几何关系,即

$$\frac{x}{X} = \frac{y}{Y} = \frac{z}{f}, \quad (3)$$

式中  $f$  为摄像机的焦距。

(3)式反映了从像点到物点之间的映射关系,其中(1)式与(2)式均为坐标系变换,摄像机的成像规律集中于(3)式中。

在摄像机坐标系下,当物点沿  $x, y, z$  中一轴向进行运动时,另外两轴向的信息保持不变,为常量。为方便讨论,当物点分别沿  $x$  轴和  $y$  轴方向移动时,取  $z = a$ , 由(3)式可得

$$\frac{x}{X} = \frac{y}{Y} = \frac{a}{f} = k_1, \quad (4)$$

也即

$$\begin{cases} x = k_1 X \\ y = k_1 Y \end{cases}, \quad (5)$$

式中  $k_1 = a/f$  为常数。

当物点沿  $z$  轴方向移动时, 设  $x = b, y = c$ , 由(3)式得

$$z = \frac{bf}{X} = \frac{cf}{Y}, \quad (6)$$

即

$$z = \frac{k_2}{X} = \frac{k_3}{Y}. \quad (7)$$

由(5)式和(7)式可以看出, 当物点分别沿  $x$  轴和  $y$  轴运动时, 成像规律成正比关系; 当物点沿  $z$  轴运动时, 成像规律成反比关系。结合(1)式和(2)式,  $(x, y, z)$  经坐标系平移与旋转变换得到  $(X_w, Y_w, Z_w)$ ,  $(X, Y)$  经坐标系平移与伸缩变换得到  $(u, v)$ , 至此, 便完成了  $(X_w, Y_w, Z_w)$  与  $(u, v)$  之间的映射关系。可知, 若旋转矩阵  $R$  为单位矩阵, 则  $X_w$  和  $Y_w$  方向的成像规律仍然保持为线性; 若  $R$  不为单位矩阵, 则坐标变换会在  $X_w$  和  $Y_w$  方向引入非线性误差; 而不管  $R$  为何值,  $Z_w$  方向的成像规律均会在反比关系的基础上变得更加复杂。

## 2.2 并行标定方法

由以上分析可知, 物点沿  $X_w$  和  $Y_w$  轴方向的成像规律接近线性, 沿  $Z_w$  轴方向的成像规律为复杂的非线性。当标定空间较小时, 上述 3 个方向成像规律差异较小, 使用相同的训练点集同时标定  $X_w, Y_w, Z_w$  信息, 能够获得较高的精度<sup>[14]</sup>; 当标定空间逐渐扩大后, 训练点集不能同时反映上述三种信息的变化规律, 导致标定精度的降低。为提高标定精度, 需要扩大标定点集的规模, 这又会导致标定速度的降低。当标定空间扩大到一定范围后, 训练精度与训练速度间的矛盾便不可调节。本文通过建立不同的训练点集得到不同的神经网络, 对  $X_w, Y_w, Z_w$  分别进行标定, 即并行标定。依据(5)式,  $X_w$  与  $Y_w$  的成像规律相似, 故可以采用相同的神经网络进行标定, 标定方式如图 3 所示。

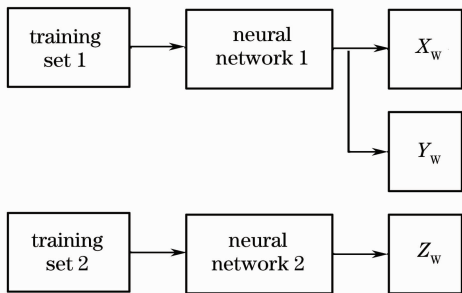


图3 并行标定方法

Fig. 3 Parallel calibration

## 2.3 归一化方法

采用反向传播(BP)神经网络, 隐层传递函数为

tansig。依据其函数特性, 需要对图像坐标进行归一化。普通归一化方法使用单一值对所有点进行归一化, 其作用仅仅是对神经网络的输入控制在传递函数的有效区间内。针对  $X_w$  和  $Y_w$  的成像规律, 提出一种新的归一化方法。

摄像机坐标系下, 物点沿  $z$  轴从位置 1 运动至位置 2, 其  $x$  和  $y$  坐标值保持不变。设两位置下标定模板中右下角角点(最大图像坐标点)  $P_m$  和其它任一角点  $P_0$  的摄像机坐标分别为  $(x_m, y_m), (x'_m, y'_m)$  和  $(x_0, y_0), (x'_0, y'_0)$ ; 相应的图像物理坐标为  $(X_m, Y_m), (X'_m, Y'_m)$  和  $(X_0, Y_0), (X'_0, Y'_0)$ 。若只考虑  $x$  轴方向, 由(5)式可得

$$\begin{cases} x_0 = k_1 X_0 \\ x_m = k_1 X_m \\ x'_0 = k'_1 X'_0 \\ x'_m = k'_1 X'_m \end{cases} \quad (8)$$

在位置 1, 使用  $P_m$  的坐标值  $X_m$  进行归一化, 则  $P_0$  的归一化坐标为

$$X = \frac{X_0}{X_m} = \frac{x_0/k_1}{x_m/k_1} = \frac{x_0}{x_m}. \quad (9)$$

在位置 2, 使用  $P_m$  的坐标值  $X'_m$  进行归一化, 则  $P_0$  的归一化坐标为

$$X' = \frac{X'_0}{X'_m} = \frac{x'_0/k'_1}{x'_m/k'_1} = \frac{x'_0}{x'_m}. \quad (10)$$

两位置中, 同一物点的  $x$  坐标不变, 即

$$\begin{cases} x_0 = x'_0 \\ x_m = x'_m \end{cases} \quad (11)$$

则有

$$X = X', \quad (12)$$

同理

$$Y = Y'. \quad (13)$$

由(12)式和(13)式可知, 同一物点  $P_0$  在两个位置的归一化坐标相同。在图像物理坐标系下, 若在模板的每个位置均使用  $P_m$  的坐标值对图像坐标进行归一化处理, 则由其中一个位置的图像坐标便可得到其他任意位置的图像坐标。但是, 实际中得到的是图像像素坐标, 模板是沿世界坐标系的  $Z_w$  轴移动, 而非  $z$  轴, 这些均涉及到坐标变换, 从而引入非线性因素; 同时, 摄像机成像中亦存在诸多非线性因素, 也导致物点在各个位置对应的图像坐标的偏移量不同。所有这些非线性因素要求构建恰当的神经网络来解决。

## 2.4 神经网络结构

BP 网络为一种多层前馈网络。由于含单隐层

的前馈网络可实现从输入到输出的任意非线性映射,因此 BP 网络可以用于拟合成像过程中的非线性映射关系。但是,一个隐层虽已足够反映这种映射关系,但并不一定最好。文献[15]证明,当样本规模为  $N$  时,双隐层前馈网络只需  $(N/2)+3$  个隐层神经元便能达到单隐层神经网络  $N-1$  个神经元所能获得的精度,即网络规模越大,双隐层网络的优越性越大。本文所进行的并行标定实验,采用的样本规模很大,故采用了双隐层神经网络;通过反复实验,确立  $5 \times 10$  的双隐层结构。

### 3 标定实验

实验中的实验设备包括:中国大恒(集团)有限公司生产的 DH-SV2000FC 工业摄像机,最大分辨率为  $1628 \text{ pixel} \times 1236 \text{ pixel}$ ,焦距  $50 \text{ mm}$ ,像素尺寸为  $4.2 \mu\text{m} \times 4.2 \mu\text{m}$ ;北京微纳光科仪器有限公司生产的 WNSC400 系列电移台控制器,电动二维平移台 WN201ZA,电动升降台 WN03VA500,电动三维旋转台 WN302RA200,手动升降台 WN03VM500,手动旋转台 WN02RM100,精密角位台 WN04GM15 等。深圳科创时代电子有限公司生产的 CG-0150S-3.00 机器视觉标定模板,模板的尺寸为  $150 \text{ mm} \times 150 \text{ mm}$ ,棋盘格的尺寸为  $3 \text{ mm} \times 3 \text{ mm}$ 。整体实验设备如图 4 所示。

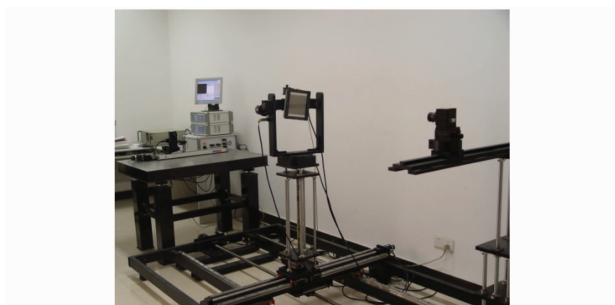


图 4 整体实验设备

Fig. 4 Photo of experimental equipment

沿世界坐标系  $Z_w$  轴正向移动标定模板,每隔  $3.125 \text{ mm}$  通过左右摄像机拍摄模板图像,总共 31 组,故整体标定空间的  $Z_w$  轴深度为  $93.75 \text{ mm}$ 。通过角点提取算法获取角点并进行匹配,每个位置获取 38 列 40 行共 1520 组角点对,如图 5 所示。标定所覆盖的空间范围为  $111 \text{ mm} \times 117 \text{ mm} \times 93.75 \text{ mm}$ ,如图 6 所示。

#### 3.1 沿 $X_w$ 和 $Y_w$ 轴方向的标定

沿  $X_w$  轴和  $Y_w$  轴同时进行标定时,采用第 1 个位置和第 31 个位置的角点对以及相应的  $X_w$  和

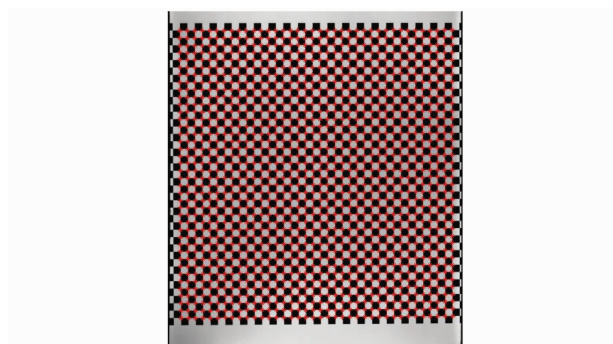


图 5 角点提取

Fig. 5 Corner extraction

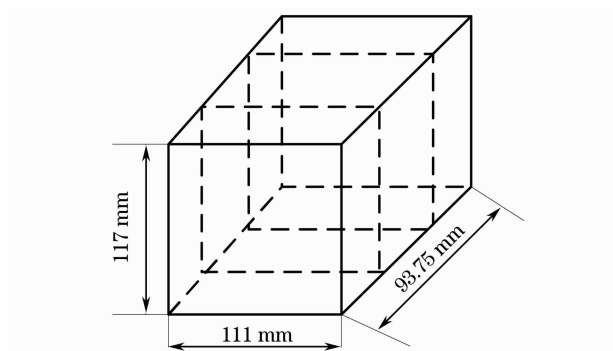


图 6 标定空间

Fig. 6 Calibration space

$Y_w$  理论值作为训练集。建立  $4 \times 5 \times 10 \times 2$  的 BP 神经网络,隐层和输出层传递函数分别为  $\text{tansig}$  和  $\text{purelin}$ ,均方误差指标为  $10^{-3}$ ,最高迭代次数为 1000。从其他位置中任选其一作为重建位置,利用已训练好的网络进行重建。

对每个位置的图像坐标集求最大值,并用其对该位置的图像坐标进行归一化。表 1 为传统归一化方法与本文方法的 5 次重构标准差对比。本文方法充分考虑了成像过程中的线性因素,为神经网络解决了部分的多余计算任务,使其更能专注于非线性部分的计算,从而提高了标定精度。

表 1 重构标准差

Table 1 Standard deviation of reconstruction

Methods	Normal method		Method in this paper	
	$\sigma_{X_w} / \text{mm}$	$\sigma_{Y_w} / \text{mm}$	$\sigma_{X_w} / \text{mm}$	$\sigma_{Y_w} / \text{mm}$
Standard deviation				
1	0.0812	0.2350	0.0569	0.0610
2	0.1191	0.0983	0.0475	0.0729
3	0.1071	0.2331	0.0594	0.0616
4	0.1204	0.1077	0.0554	0.0563
5	0.1727	0.1045	0.0695	0.0551
Average value	0.1201	0.1557	0.0577	0.0613

### 3.2 沿 $Z_w$ 轴方向标定

保持重构位置不变。剩余 30 个位置包括了 1520 个点沿  $Z_w$  轴向运动的 30 个位置的信息, 本文从 1520 个点中选取了 81 个点的  $Z_w$  信息(共 2430 角点对)作为训练点集, 其分布如图 7 所示。建立  $4 \times 5 \times 10 \times 1$  的 BP 神经网络, 其他参数设置与上节相同。对重建位置进行  $Z_w$  信息重构, 结果如图 8 所示。

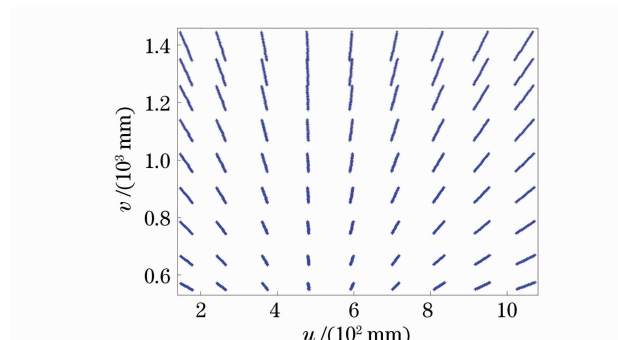


图 7 训练点集分布

Fig. 7 Distribution of training set

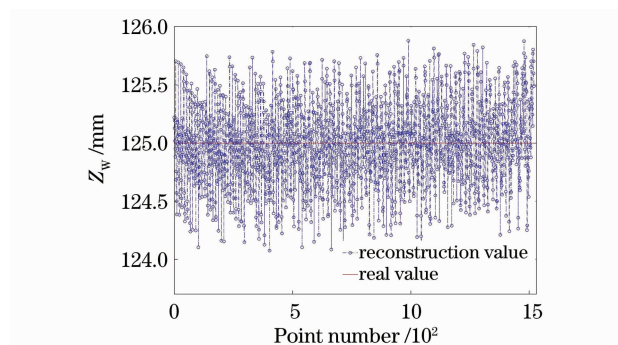


图 8  $Z_w$  重构结果

Fig. 8 Reconstruction of  $Z_w$  coordinates

经计算, 该位置 1520 个空间点的  $Z_w$  重构标准差为 0.3674 mm。可以看出,  $Z_w$  的标准差要远远大于  $X_w$  和  $Y_w$  的标准差。这表明, 影响总体标定精度的主要因素为  $Z_w$  的标定精度。  $Z_w$  方向的成像规律比  $X_w$  和  $Y_w$  的成像规律复杂得多; 同时  $Z_w$  成像规律比较复杂, 均匀分布的训练点集无法充分反映其规律, 导致标定精度降低。

### 3.3 总体实验分析

根据误差合成, 并行标定方法的三维重建标准差计算方法为

$$\delta = \sqrt{\delta_{X_w}^2 + \delta_{Y_w}^2 + \delta_{Z_w}^2}, \quad (14)$$

结合上述实验数据, 计算得到最终的三维重建标准差为 0.3769 mm。文献[16]使用一个训练集对三维信息同时进行标定, 虽然其三维重建精度达到了

$10^{-3}$  量级, 但前提是对输出也进行归一化, 而且其归一化基数并未提到, 但取值范围应该在几百到几毫米。因此, 本文方法能够得到与文献[16]相当的三维重建精度。

同时, 沿  $X_w$  和  $Y_w$  轴向进行标定时, 训练点集规模为 3040 对应点; 沿  $Z_w$  轴方向进行标定时, 训练点集的规模为 2430 对应点; 而标定空间中包含 47120 对应点。故训练集规模仅占整个空间的 11.6%。训练点集的规模直接影响神经网络的学习速度, 因此, 本文能够在保证标定速度的前提下扩大标定空间。

## 4 结 论

提出了一种新的并行标定方法和一种新的归一化方法, 扩大了标定空间, 并进一步提高了  $X_w$  轴和  $Y_w$  轴向的标定精度。实验证明,  $Z_w$  轴方向的标定为整体大范围标定工作的重点, 为后续研究指明方向。研究表明, 只有充分遵循摄像机成像规律, 才能合理训练点集, 更好发挥神经网络的特点, 进而在保证标定精度和标定速度的同时, 扩展标定空间, 提高标定方法的实用性。

## 参 考 文 献

- 1 Ge Baozhen, Li Xiaojie, Qiu Shi. Camera lens distortion correction based on coplanar point direct linear transformation [J]. *Chinese J. Lasers*, 2010, **37**(2): 488~494  
葛宝臻, 李晓洁, 邱实. 基于共面点直接线性变换的摄像机畸变校正[J]. *中国激光*, 2010, **37**(2): 488~494
- 2 R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-shelf TV cameras and lenses[J]. *IEEE J. Robotics and Automation*, 1987, **3**(4): 322~344
- 3 Mao Xianfu, Su Xianyu, Liu Yuankun *et al.*. Analysis on optical coordinate measurement based on phase target[J]. *Acta Optica Sinica*, 2009, **29**(9): 2452~2457  
毛先富, 苏显渝, 刘元坤等. 基于相位标靶的光学坐标测量方法[J]. *光学学报*, 2009, **29**(9): 2452~2457
- 4 Sun Junhua, Liu Zhen, Zhang Guangjun *et al.*. Camera calibration based on flexible 3D target[J]. *Acta Optica Sinica*, 2009, **29**(12): 3433~3439  
孙军华, 刘震, 张广军等. 基于柔性立体靶标的摄像机标定[J]. *光学学报*, 2009, **29**(12): 3433~3439
- 5 Huo Ju, Yang Wei, Yang Ming. A self-calibration technique based on the geometry property of the vanish point[J]. *Acta Optica Sinica*, 2010, **30**(2): 465~472  
霍炬, 杨卫, 杨明. 基于消失点几何特性的摄像机自标定方法[J]. *光学学报*, 2010, **30**(2): 465~472
- 6 Zhu Jia, Li Xingfei, Xu Yingxin. Camera calibration technique based on active vision[J]. *Acta Optica Sinica*, 2010, **30**(5): 1297~1303  
朱嘉, 李醒飞, 徐颖欣. 摄像机的一种主动视觉标定方法[J]. *光学学报*, 2010, **30**(5): 1297~1303
- 7 J. J. Hopfield, D. W. Tank. Neural networks and physical systems with emergent collective computational abilities[J].

- Proceedings of the National Academy of Sciences*, 1982, **79**(8): 2554~2558
- 8 D. E. Rumelhart, J. L. McClelland. Parallel Distributed Processing[M]. Cambridge: MIT Press, 1986
- 9 J. Wen, G. Schweitzer. Hybrid Calibration of CCD cameras using artificial neural nets[C]. Proc. Int Joint Conf. Neural Networks (New York), 1991, **1**: 337~342
- 10 M. Lynch, C. Dagi. Backpropagation neural network for stereoscopic vision calibration[C]. *SPIE*, 1992, **1615**: 289~298
- 11 J. Jun, C. Kim. Robust camera calibration using neural network [C]. Proceedings of the IEEE Region 10 conference, 1999, 694~697
- 12 Liu Hongjian, Luo Yi, Liu Yuncai. Variable precision camera calibration using neural network [J]. *Optics and Precision Engineering*, 2004, **12**(4): 443~448  
刘宏建, 罗毅, 刘允才. 可变精度的神经网络摄像机标定法 [J]. *光学精密工程*, 2004, **12**(4): 443~448
- 13 Zhang Guangjun. Machine Vision[M]. Beijing: Science Press, 2005. 24~27  
张广军. 机器视觉[M]. 北京: 科学出版社, 2005. 24~27
- 14 Liu Sheng, Fu Huixuan, Wang Yuchao. Camera calibration for stereo vision based on LS-SVM [J]. *Opto-Electronic Engineering*, 2008, **35**(10): 21~25  
刘胜, 傅荟璇, 王宇超. 基于LS-SVM的立体视觉摄像机标定 [J]. *光电工程*, 2008, **35**(10): 21~25
- 15 Shin'ichi Tamura, Masahiko Tateishi. Capabilities of a four-layered feedforward neural network; four layers versus three[J]. *IEEE Transactions on Neural Networks*, 1997, **8**(2): 251~255
- 16 Yuan Ye, Ou Zongying, Tian Zhongxu. 3D reconstruction of stereo vision using neural networks implicit vision model[J]. *J. Computer-Aided Design & Computer Graphics*, 2003, **15**(3): 293~296  
袁野, 欧宗瑛, 田中旭. 应用神经网络隐式视觉模型进行立体视觉的三维重建[J]. *计算机辅助设计与图形学学报*, 2003, **15**(3): 293~296