

## 基于深度学习的散斑投影轮廓术

钟锦鑫<sup>1,2</sup>, 尹 维<sup>1,2</sup>, 冯世杰<sup>1,2</sup>, 陈 钱<sup>2</sup>, 左 超<sup>1,2</sup>

- (1. 南京理工大学 电子工程与光电技术学院 智能计算成像实验室 (SCILab), 江苏南京 210094;
2. 南京理工大学 江苏省光谱成像与智能感知重点实验室, 江苏南京 210094)

**摘 要:** 针对传统的单幅散斑图像匹配算法测量精度低且无法测量复杂面型物体等问题, 提出了一种基于深度学习的散斑投影轮廓术, 即通过深度学习的方法实现散斑图像的逐像素匹配。设计利用孪生卷积神经网络结构, 将目标散斑图像和参考散斑图像以图像块的形式输入神经网络。通过卷积层运算提取散斑图像块的特征信息, 进而将子网络得到的特征信息融合为两个图像块之间的匹配系数, 以获得散斑图像的视差数据, 并最终可将视差数据转化为物体的三维信息。实验结果表明, 该方法可以通过单幅散斑图像实现精度约为 290  $\mu\text{m}$  的三维轮廓测量。

**关键词:** 散斑投影; 深度学习; 三维测量; 孪生网络

**中图分类号:** TB133; TP391 **文献标志码:** A **DOI:** 10.3788/IRLA20200011

## Speckle projection profilometry with deep learning

Zhong Jinxin<sup>1,2</sup>, Yin Wei<sup>1,2</sup>, Feng Shijie<sup>1,2</sup>, Chen Qian<sup>2</sup>, Zuo Chao<sup>1,2</sup>

- (1. Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;
2. Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** Traditional single speckle pattern matching algorithms always suffer from the low measurement accuracy and cannot be used to measure complex surface objects. A speckle projection profilometry with deep learning was proposed to realize the pixel-by-pixel matching. The siamese convolutional neural network structure was applied and extended where the main speckle pattern and the auxiliary speckle pattern were fed into the neural network patch by patch. It was expected that the feature from the speckle pattern patches could be extracted by the convolution operation. In this way, the features were fused and the matching coefficient between the two patches was obtained, which could be further used to formulate the disparity data and then the three-dimensional (3D) object was reconstructed. The experiment results demonstrate that with the proposed method 3D measurement with an accuracy of about 290  $\mu\text{m}$  could be achieved through a single speckle pattern.

**Key words:** speckle projection; deep learning; 3D measurement; siamese network

收稿日期: 2020-03-01; 修订日期: 2020-04-10

基金项目: 国家自然科学基金 (61722506, 61705105, 11574152); 江苏省杰出青年基金 (BK20170034); 江苏省重点研发计划 (BE2017162); 江苏省光谱成像与智能感知重点实验室开放基金 (3091801410411)

作者简介: 钟锦鑫 (1996-), 男, 硕士生, 主要从事条纹投影、深度学习等方面的研究。Email: jinxzleo@163.com

导师简介: 左超 (1987-), 男, 教授, 博士生导师, 博士, 主要从事计算光学显微成像、快速三维传感等方面的研究。

Email: zuochao@njjust.edu.cn (通讯联系人)

## 0 引言

近年来,三维形状测量技术在医学成像、机器人导航、人脸识别、工业检测和人机交互等领域有着重要的地位<sup>[1-4]</sup>。双目立体视觉系统作为一种经典的三维测量系统,因其易于实现而被广泛采用<sup>[5]</sup>。利用左右两个相机获取物体在不同视角的图像,通过计算两幅图像对应点之间的位置偏差,即视差数据,获得物体的三维信息。因此如何准确地求解物体在两个视角的匹配关系是恢复物体三维信息的关键。基于结构光的双目立体视觉系统是通过投影按照一定模式和规则编码的图像到待测物体上,并使用相对应的解码策略获得高精度的视差数据。

最为常见的两种结构光三维测量方法分别为:条纹投影和散斑投影。其中条纹投影是根据形变的条纹图像计算出待测物体在两个视角的相位信息,利用相位信息获得待测物体在两个视角的匹配信息。虽然通过相位匹配可以获得准确的视差数据,但是由于正弦条纹本身的周期性特性,使其无法通过单幅条纹图像获得准确且连续的相位信息,这在一定程度上限制了条纹投影在动态测量中的应用。散斑投影是通过投影随机散斑图像到物体表面,增加物体表面的特征信息。利用相似函数对散斑图像进行计算,获得两个视角的散斑图像之间的匹配关系。虽然散斑投影可以通过单幅散斑图像实现三维测量,但是利用传统的相似函数,例如零均值归一化互相关函数(ZNCC)<sup>[6]</sup>和零均值归一化差平方和函数(ZNSSD)<sup>[7]</sup>等,获得的视差数据精度往往远低于通过相位匹配获得的视差数据。

目前,传统的散斑投影测量方法存着许多的不足之处,例如测量速度慢、测量精度低等。针对散斑投影测量速度慢的问题,可以通过 GPU 加速<sup>[8]</sup>和利用激光散斑<sup>[9]</sup>实现快速投影的方法提高散斑投影的测量速度。针对散斑投影测量精度低的问题,可以改变投影的散斑图案,例如使用不同密度的散斑图案<sup>[10]</sup>、RGB 散斑图案<sup>[11]</sup>等,或使用更准确的匹配算法,例如利用不同散斑图像实现的去相关匹配算法<sup>[12]</sup>、利用四幅不同密度的二值散斑图像实现的时空逻辑相关(STL)立体匹配算法<sup>[13]</sup>等,提高散斑投影的测量精度。虽然新的匹配算法可以实现高精度的散斑投影

测量,但是仍然无法通过单幅散斑图像获得准确的三维数据。

与传统方法相比,深度学习具有强大的特征学习和特征表达能力,使其可以解决多个传统方法难以解决的问题。伴随着计算能力的不断提升,深度学习已经在多个领域,计算机视觉、语音识别等,取得了出色的成果<sup>[14]</sup>。在计算机视觉领域,基于卷积神经网络(CNN)的各种深度学习网络框架解决了计算成像的诸多难题,例如相位检索<sup>[15]</sup>、无透镜显微成像<sup>[16]</sup>、条纹分析<sup>[17]</sup>、高速测量<sup>[18]</sup>等。文中提出了基于深度学习的散斑投影轮廓术。采用双目立体视觉系统,将散斑图像作为神经网络的输入,并通过条纹投影轮廓术计算得到的物体在两个视角的匹配关系作为神经网络的真实值,对神经网络的训练数据进行标记。文中所使用的神经网络结构为基于卷积神经网络的孪生网络结构,借助卷积神经网络对图像数据高级抽象特征提取的优势,将散斑图像转换为特征数据,同时孪生网络可以实现左右视角散斑图像的权重共享。以该网络结构对训练数据进行多次训练,利用训练好的深度学习模型对待测物体在左右视角采集的单幅散斑图像进行特征提取,并且通过散斑图像的特征数据计算匹配代价聚合,从而获得散斑图像的视差数据。最终,结合双目立体视觉系统的标定参数,可将视差数据转换为待测物体的三维信息。实验证明,文中所提出的基于深度学习的方法可以通过单幅散斑图像获得准确的视差数据和物体三维信息。

## 1 基于相位辅助的高精度视差标记

文中使用的深度学习方法为监督学习,因此在训练时需要为输入数据提供与其相对应的真实值对输入的训练数据进行标记。神经网络的输入数据为散斑图像,其真实值为物体在两个视角间的匹配关系。为了获得准确的匹配关系,采用条纹投影轮廓术计算训练数据集中物体在两个视角的视差数据。

利用投影仪投影正弦条纹,通过两个视角的相机对投影到待测物体上的条纹图像进行采集。投影到待测物体上的条纹会根据待测物体形状的不同产生不同程度的形变。条纹投影轮廓术就是通过这些形变条纹求解出物体在两个视角上的相位信息,并利用相位信息进行匹配。目前,有许多的方法可以从形变

条纹中提取出相位信息,其中  $N$  步相移法可以提供较高的测量分辨率和精度。投影仪投影  $N$  步相移法所使用的正弦条纹  $I_n^p(x^p, y^p)$  可以表示为:

$$I_n^p(x^p, y^p) = \frac{1}{2} + \frac{1}{2} \cos\left(2\pi f_0 x^p - \frac{2\pi n}{N}\right) \quad (1)$$

式中:  $f_0$  为正弦条纹的频率;  $(x^p, y^p)$  为投影仪坐标系中的像素坐标。在投影仪投影正弦条纹到待测物体上后,通过相机来采集投影到待测物体上的正弦条纹,采集到的正弦条纹  $I_n(x, y)$  可以表示为:

$$I_n(x, y) = A(x, y) + B(x, y) \cos\left[\phi(x, y) - \frac{2\pi n}{N}\right] \quad (2)$$

式中:  $(x, y)$  为相机坐标系中的像素坐标;  $A(x, y)$  为平均光强;  $B(x, y)$  为光强调制度;  $\phi(x, y)$  为待测物体的相位<sup>[19]</sup>。通过最小二乘法<sup>[20]</sup>可以将相位  $\phi(x, y)$  表示为:

$$\phi(x, y) = \arctan \frac{\sum_{n=1}^N I_n \sin\left(\frac{2\pi n}{N}\right)}{\sum_{n=1}^N I_n \cos\left(\frac{2\pi n}{N}\right)} \quad (3)$$

根据反正切函数的特性,计算出来的相位  $\phi(x, y)$  的数值范围为  $-\pi$  到  $\pi$ , 并且存在着  $2\pi$  的不连续跳变。将公式 (3) 中计算出来的相位  $\phi(x, y)$  称为包裹相位。由于包裹相位具有周期性特性,因此在相机的同一极线上存在着多个相位相同的像素点,所以无法通过包裹相位来进行相位匹配。为了消除包裹相位中的不连续跳变,从而获得连续分布的相位信息,需要通过相位展开算法将包裹相位  $\phi(x, y)$  转换为绝对相位  $\Phi(x, y)$ 。其中绝对相位和包裹相位之间的关系可以表示为:

$$\Phi(x, y) = \phi(x, y) + 2\pi k(x, y) \quad (4)$$

式中:  $k(x, y)$  为光栅条纹级次。根据获取光栅条纹级次  $k(x, y)$  的原理不同,可以将相位展开算法分为空间展开算法和时间展开算法。空间展开算法是利用包裹相位与其相邻像素的相位之间的关系来进行相位展开<sup>[21]</sup>。对于相邻像素中相位差较大的不连续表面,空间相位展开算法无法准确地消除包裹相位中的不连续跳变。因为空间相位展开算法在处理不连续表面中存在的缺陷,所以提出了多频时间相位展开算法。

多频时间相位展开算法通过多个不同频率的包裹相位来进行相位展开<sup>[22]</sup>,通过投影不同频率的条纹图像到待测物体表面,计算出不同频率的包裹相位:

高频包裹相位  $\phi_h(x, y)$  和低频包裹相位  $\phi_l(x, y)$ , 与这两个频率的包裹相位相对应的绝对相位可以表示为:

$$\Phi_h(x, y) = \frac{f_h}{f_l} \Phi_l(x, y) \quad (5)$$

式中:  $\Phi_h(x, y)$  为高频包裹相位  $\phi_h(x, y)$  所对应的绝对相位;  $\Phi_l(x, y)$  为低频包裹相位  $\phi_l(x, y)$  所对应的绝对相位;  $f_h$  和  $f_l$  分别为高频和低频条纹的条纹频率。根据公式 (4) 和公式 (5) 可以将高频条纹所对应的光栅条纹级次  $k_h(x, y)$  表示为:

$$k_h(x, y) = \text{Round}\left[\frac{\phi_l(x, y)f_h/f_l - \phi_h(x, y)}{2\pi}\right] \quad (6)$$

式中:  $\text{Round}[\ ]$  为取整函数。在已知光栅条纹级次和包裹相位的情况下,就可以根据公式 (4) 计算得到与包裹相位相对应的绝对相位。条纹投影轮廓术获得物体在两个视图之间的视差信息的步骤为:首先,通过两个视角的相机采集投影到物体上的正弦条纹图像,然后,利用  $N$  步相移法将条纹图像转化为包裹相位,再利用时间相位展开算法对包裹相位进行计算得到绝对相位,最后,通过绝对相位匹配获得物体在两个视图之间的视差信息。

相位匹配是在极线上通过最小化绝对相位差获得物体在两个视图之间的视差信息。以目标图像的像素点  $(a, b)$  为例,在参考图像的同一极线上搜索与点  $(a, b)$  绝对相位误差最小的点  $(c, b)$ , 两个点在横坐标的差  $a-c$  就为目标图像中像素点  $(a, b)$  的视差值。为了确保相位匹配的准确性,需要保证两个视图的绝对相位处于同一极线上。通过对双目立体视觉系统进行标定,利用标定数据对采集到的条纹图像进行极线校正,使得两个视图的条纹图像处于同一极线上,从而保证获得的绝对相位处于同一极线。

## 2 深度学习神经网络结构

深度学习的方法是利用神经网络结构实现对数据的分析和预测。在该章节中介绍的神经网络结构是基于卷积神经网络建立的孪生网络结构。孪生网络由两个子网络组成,两个子网络通过共享权重实现连接。神经网络结构如图 1 所示,以目标散斑图像中的一个像素点做边长为  $R$  的正方形图像块,以参考散斑图像中  $D$  个参考像素点做长为  $R+D-1$ , 宽为  $R$  的长方形图像块。在文中,目标散斑图像块的大小为

19×19, 即  $R=19$ 。根据两个相机之间的基线距离和对训练集中视差数据的统计, 训练集物体的视差数据范围为-50~100, 所以参考散斑图像块的大小为 19×169, 即  $D=151$ 。将这两个图像块作为子网络的输入数据, 两个子网络的基本结构由卷积层组成。卷积层的作

用是提取输入数据的特征值, 因为卷积层中卷积核的尺寸为 3, 且采用的 padding 方式为 VALID, 所以经过 9 卷积层后输出的特征矩阵分别为  $[1,64]$  和  $[151,64]$ , 其中 64 为卷积层中卷积核个数即输出的特征数。

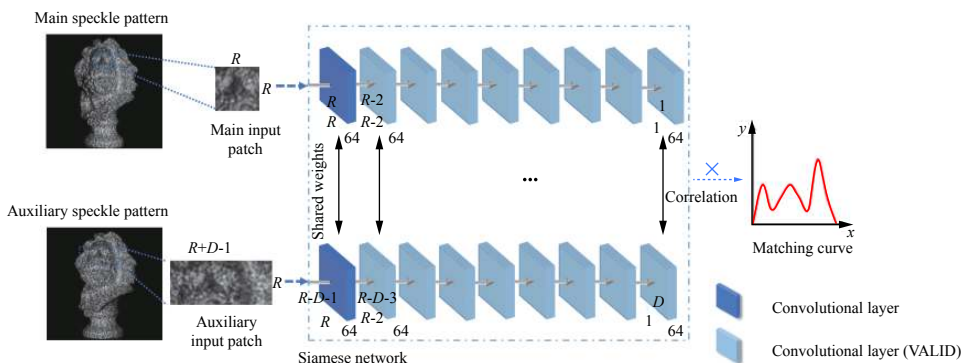


图 1 神经网络结构

Fig.1 Neural network architecture

在输入数据经权重共享的子网络处理后, 目标像素点与参考像素点所对应的特征矩阵将经过点积运算获得匹配相似度。深度学习中需要对训练数据进行标记, 因为该神经网络的作用是在参考散斑图像的 151 个参考像素点中寻找目标散斑图像目标像素点的匹配点, 所以神经网络的“标签”应为目标散斑图像的目标像素点与参考散斑图像的 151 个参考像素点的匹配相似度。将匹配相似度以曲线图的形式表示, 匹配曲线如图 2 所示。匹配曲线的取值可以表示为:

$$y_i = \begin{cases} 0.5, & x_i = x_c \\ 0.1, & abs(x_i - x_c) = 1 \\ 0.05, & abs(x_i - x_c) = 2 \\ 0, & \text{其他} \end{cases} \quad (7)$$

式中:  $y_i$  为曲线纵坐标, 即匹配相似度;  $x_i$  为曲线横坐标, 若目标散斑图像的目标点为  $(a, b)$ , 参考散斑图像的匹配点为  $(c, b)$ , 则目标点的视差值为  $a-c$ 。因为 151 个参考像素点在横坐标分布范围为:  $a-100 \rightarrow a+50$ , 所以目标点  $(a, b)$  在曲线图横坐标轴上对应的数值为  $x_a=100$ , 匹配点  $(c, b)$  在曲线图横坐标轴上对应的数值为  $x_c=100+c-a$ 。因此匹配曲线峰值所对应的横坐标会随着目标点的视差值而变化。

相比于传统的单输入卷积神经网络<sup>[23]</sup>, 文中的网

络采用了孪生网络结构, 由于孪生网络利用权重共享的子网络来统一处理不同路径的输入数据, 因此可以更好地衡量两个输入图像的相似度, 所以孪生网络结构更加适用于图像匹配。相比于 Zbontar J 的 MC-CNN 网络<sup>[24]</sup>, 文中的网络没有正样本与负样本选择的局限性, 在参考图像块中将所有可能的匹配点作为参考像素点, 并且在输出端使用较为平缓的真实值曲线容许了一定的误差范围, 从而使网络的匹配结果更稳定、可靠。相比于 Luo W<sup>[25]</sup> 的网络结构, 因为 Luo W 的网络结构中参考图像块将匹配点  $(c, b)$  置于图像块中心, 即参考像素点的横坐标分布为:  $c-75 \rightarrow c+75$ , 所以真实值曲线中最高点的横坐标始终固定在  $x_i=75$  的位置。而文中的网络结构中真实值的曲线最

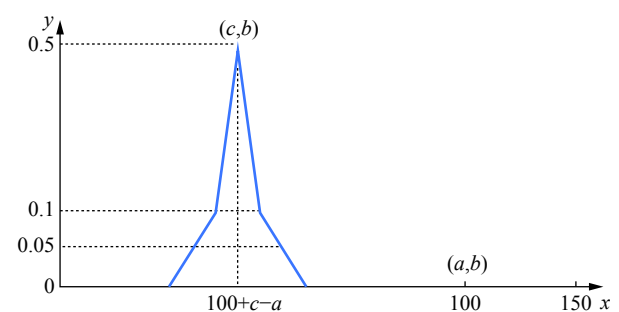


图 2 神经网络“标签”

Fig.2 Label of neural network



高点会随着输入目标图像块视差值的不同而变化,这样增加了神经网络训练时的不确定性,从而提高了网络的稳定性和预测的准确性。

### 3 基于深度神经网络的散斑位置预测

通过上述网络结构对输入数据进行训练,并利用训练好的模型对散斑图像进行预测。与训练时输入数据为图像块不同,为了减少预测的时间和使用资源,在使用深度学习模型预测时输入的测试数据为整幅图像。通过深度学习模型对目标散斑图像和参考散斑图像进行预测,其中目标散斑图像和参考散斑图像分别为左右视角相机采集的单幅散斑图像。利用子网络预测获得的特征值进行计算,获得散斑图像的视差匹配值,即匹配代价聚合 $C(x,y,d)$ ,匹配代价聚合的计算公式如下:

$$C(x,y,d) = -F_m(x,y,f) \cdot F_r(x+d-100,y,f) \quad (8)$$

式中: $F_m$ 为目标散斑图像预测得到的特征值; $F_r$ 为参考散斑图像预测得到的特征值; $f$ 为卷积层输出的特征数; $d$ 为视差值,因为训练时使用的参考散斑图像由 151 个参考像素点组成,所以 $d \in [0,150]$ ; $(x,y)$ 为目标散斑图像的像素点坐标。

尽管通过神经网络可以得到较为准确的匹配结果,但是还不足以获得准确的视差信息,在一些边缘和特征复杂的区域会出现较多的错误点,所以可以对匹配代价聚合 $C(x,y,d)$ 进行一些后处理方法来减少这些区域内的错误点,从而获得更准确的视差数据。因为在后处理中使用到半全局匹配算法(SGM),所以为了适应半全局匹配算法,在公式(8)中将视差匹配代价聚合以负数形式表示,使得匹配点的视差匹配值为视差匹配代价聚合中的最小值。

#### 3.1 半全局匹配

将从子网络中预测获得的视差匹配代价聚合通过半全局匹配(SGM)算法来对其进行改善,进而获得更准确的视差匹配代价聚合 $C_{SGM}(p,d)$ ,与Hirschmuller<sup>[26]</sup>中使用 16 个方向进行半全局匹配不同,文中只在水平和垂直两个方向上进行半全局匹配。目标像素点在指定方向 $r$ 上进行半全局匹配后得到的匹配代价聚合 $L_r(p,d)$ 可以表示为:

$$L_r(p,d) = C(p,d) - \min_k L_r(p-r,k) + \min \left\{ \begin{array}{l} L_r(p-r,d), \\ L_r(p-r,d-1) + P_1, \\ L_r(p-r,d+1) + P_1, \\ \min_i L_r(p-r,i) + P_2 \end{array} \right\} \quad (9)$$

式中: $p$ 为目标像素点,其坐标为 $(x,y)$ ; $P_1$ 和 $P_2$ 为惩罚系数。因为与主像素点进行半全局匹配的参考点为与主像素点在水平和垂直方向相邻的四个参考像素点,所以最终半全局匹配算法计算得到的视差匹配代价聚合 $C_{SGM}(p,d)$ 可以表示为:

$$C_{SGM}(p,d) = \frac{1}{4} \sum_r L_r(p,d) \quad (10)$$

通过计算每个目标像素点 $p$ 的匹配代价聚合 $C_{SGM}(p,d)$ 最小值所对应的视差值 $d$ 就可以得到视差数据 $D_{int}(p)$ ,视差数据可以表示为:

$$D_{int}(p) = \arg \min_d C_{SGM}(p,d) \quad (11)$$

#### 3.2 亚像素计算

根据公式(11)可知,计算得到的视差值为整数。在实际情况中视差数据基本不会以整数的形式存在。整像素的视差数据用于计算待测物体的三维数据的精度会低于亚像素的视差数据。为了可以进一步提高测量的精度,可以通过视差匹配代价聚合 $C_{SGM}(p,d)$ 拟合的曲线将整像素的视差数据转化为亚像素的视差数据:

$$D_s(p) = D_{int}(p) - \frac{C_{SGM}(p,d'+1) - C_{SGM}(p,d'-1)}{2(C_{SGM}(p,d'+1) + C_{SGM}(p,d'-1) - 2C_{SGM}(p,d'))} \quad (12)$$

式中: $D_s(p)$ 为亚像素的视差数据; $d'$ 为 $p$ 点的整像素视差数据,即 $d'=D_{int}(p)$ 。因为匹配获得的视差数据是目标图像像素点的横坐标值减去参考图像匹配点的横坐标值,所以亚像素的视差数据 $D_s(p)$ 需要根据公式(8)中的相对关系进行转换:

$$D(p) = 100 - D_s(p) \quad (13)$$

式中: $D(p)$ 为经过转换后的亚像素视差数据。虽然半全局匹配和亚像素计算的后处理方法可以获得更准确的视差数据,但是计算出的视差数据中依然会存在着部分误差点。

#### 3.3 左右一致性检验

为了提高测量的精度,需要尽可能地排除视差数

据中的误差点。在双目立体视觉系统中,可以通过左右一致性检验的方法来排除视差图像中的误差点。将左右相机采集到的散斑图像互为目标图像和参考图像,进而获得左右相机的视差图像,之后检验两个视差图像中对应像素点的视差值是否相同来判定该像素点是否为误差点,左右一致性检验的算法为:

$$D_{LR}(p) = \begin{cases} d, |d - D_R(p-d)| \leq 1 \\ 0, |d - D_R(p-d)| > 1 \end{cases} \quad (14)$$

式中:  $D_{LR}(p)$  为经过左右一致性检验后的视差图像;  $D_R(p)$  为右相机的视差图像;  $d$  为左相机视差图像  $p$  点处的视差值  $d=D_L(p)$ 。通过左右一致性检验可以将左右相机的视差图像中不对应的像素点视为误差点,然后将其去除。

### 4 实验

该实验是通过深度学习的方法来实现散斑投影轮廓术。其中深度学习的训练数据、验证数据和测试数据都需要通过双目立体视觉系统采集。双目立体视觉系统由两个相机 (Basler acA640 750  $\mu\text{m}$ ) 和一个投影仪 (DLP 4500pro) 组成。投影仪同时投影条纹图像和散斑图像,因为条纹投影的方法中计算绝对相位需要使用  $N$  步相移法和多频时间相位展开算法,所以投影仪需要投影多幅条纹图像。在该实验中,多频时间相位展开算法使用的条纹频率分别为: 1、8、64,与这三个频率的条纹相对应的条纹相移步数分别为: 3、3、9,具体条纹频率和相移步数的选取可见参考文献 [27]。两个视角的相机同时采集物体的条纹图像和散斑图像,相机获取的图像如图 3 所示,其中图 3(a) 为待测物体,图 3(b) 为左相机采集到的高频条纹图像即 64 频率的条纹图像,图 3(c) 为左相机采集到的散斑图像,图 3(d) 为右相机采集到的散斑图像。

利用双目立体视觉系统对多个物体在多个场景下进行图像采集,将左相机采集到的图像作为目标图像,右相机采集到的图像作为参考图像,共采集 750 组图像,其中 350 组为训练图像,200 组为验证图像,200 组为测试图像。一组图像中包含左右视角各 15 幅条纹图像和 1 幅散斑图像,共 32 幅图像。对训练图像中的条纹图像通过  $N$  步相移法和多频时间相位展开算法获得 350 组物体的绝对相位并且通过绝对相位匹配获得物体的视差数据,左相机采集到的训

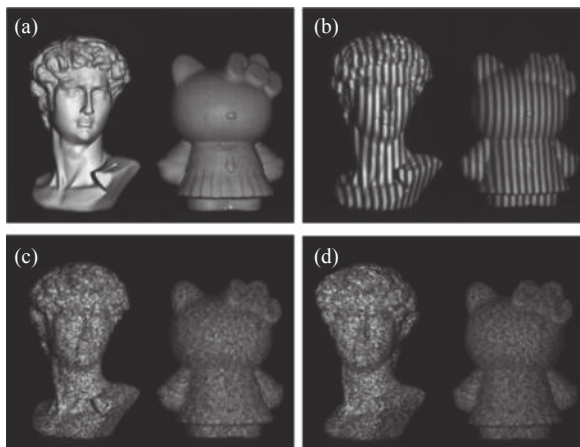


图 3 双目立体系统采集数据; (a) 待测物体图像; (b) 高频条纹图像; (c) 左相机采集的散斑图像; (d) 右相机采集的散斑图像  
Fig.3 Image data captured by stereo system. (a) Measured object pattern; (b) high-frequency fringe pattern; (c) speckle pattern from left camera; (d) speckle pattern from right camera

练数据如图 4 所示,其中包括了三个频率中零相位的条纹图像、左相机条纹图像计算获得的绝对相位和左视角的散斑图像。在该实验中条纹投影的方法是为了标记训练数据。将 350 组图像中的散斑图像以图像块的方式作为神经网络的输入数据。值得注意的是虽然训练图像只有 350 组,但是神经网络的训练数据为以有效像素点为中心的图像块,一组散斑图像中存在着成千上万个有效像素点,所以 350 组训练图像中存在着上百万组训练数据。已知有效像素点的视差数据,则可以通过公式 (7) 获得训练数据的匹配曲线,即神经网络的“标签”,对训练数据进行标记,然后通过监督学习的方式对输入的训练数据进行训练。最后利用训练完成的深度学习模型对测试物体的散斑图像进行预测,并经过后处理算法,获得待测物体的视差数据。

以其中一组测试散斑图像为例,同样使用左右视角的单幅散斑图像,将文中的方法获得的实验结果与传统的匹配算法和基于端到端的 CNN 网络获得的结果进行对比,对比结果如图 5 所示。通过传统的 ZNCC 匹配算法对目标散斑图像和参考散斑图像进行逐像素匹配,ZNCC 匹配获得待测物体的视差图像如图 5(a) 所示。以端到端形式搭建的串联卷积网络对训练数据进行训练,利用训练完成的深度学习模型对目标散斑图像和参考散斑图像进行预测,预测的结

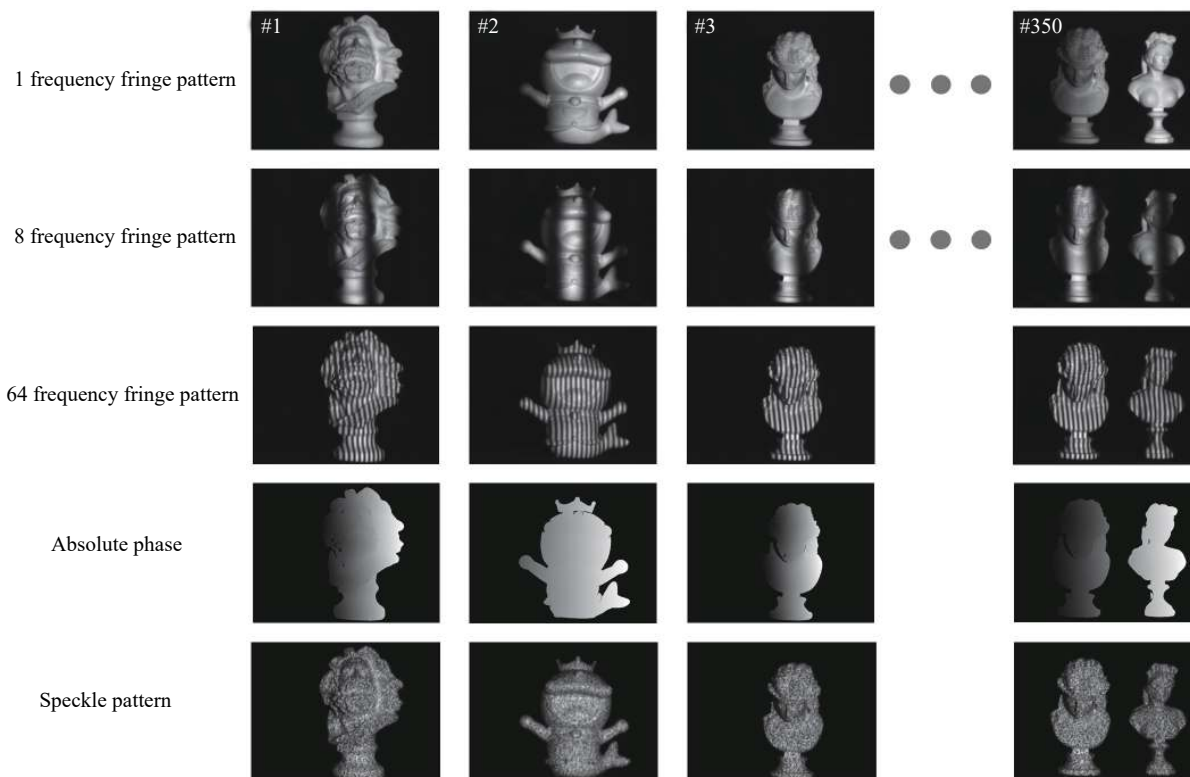


图 4 左相机采集到的训练数据和通过条纹图像计算获得的绝对相位

Fig.4 Training data captured by left camera and the absolute phase calculated by fringe patterns

果如图 5(b) 所示。使用文中所提出的方法的结果如图 5(c) 所示。为了定量地比较三种方法得到的视差图像,以条纹投影轮廓术获得的视差数据作为真值,计算三种方法获得的视差数据的误差。三种方法的误差分别如图 5(d), 5(e) 和 5(f) 所示。从误差图中可以看出,传统的 ZNCC 逐像素匹配算法对于右边面型较为简单的物体误差较小,而对于左边面型较为复杂物体就难以匹配得到正确的视差值。单输入的 CNN 网络可以预测得到待测物体大部分的视差数据,但是对于物体细节的部分,例如五官和头发等位置就无法得到准确的视差值。相比于前两种方法,文中的方法可以得到更准确的视差数据,但是在待测物体边缘的位置仍存在着少量误差点。

为了实现物体的三维测量,需要将视差数据转化为三维数据。通过对双目立体视觉系统进行标定,获得双目立体视觉系统的标定数据。以标定参考面作为三维坐标系的零平面,利用标定数据将视差数据转换为深度数据,上述三种方法获得的视差数据所对应的深度数据分别如图 5(g), 5(h) 和 5(i) 所示。从深度

图中可以更好地比较三种方法的测量精度。其中 ZNCC 匹配获得的深度图像在复杂面型的物体上存在着多处缺失。单输入 CNN 网络获得的深度图中只能看到物体大致的形状,无法获得细节处深度数据。而文中的方法获得的深度数据可以获得更多精细的轮廓。通过比较测量结果可以看出传统 ZNCC 匹配方法测量精度最差, CNN 网络结构的测量精度也比较粗糙。因为 CNN 网络是以端到端的方式对散斑图像进行训练,这使得神经网络无法顾及所有的有效点,所以只能获得物体大致的形状。相比较于 CNN 网络,因为文中的方法是以逐像素的方法进行训练,所以在被测物体的细节处仍可以获得不错的视差数据。因为在边缘处像素点的图像块中大部分区域为背景图像,所以基于局部区域匹配的方法会存在边缘处匹配不准确的问题。经实验结果表明,在同样只使用一张散斑图像获得物体的三维信息的情况下,文中的方法可以获得最好的物体三维测量结果。

为了体现基于深度学习的散斑投影轮廓术的鲁棒性和稳定性,笔者等对多个形状和大小不同的物体



进行了三维测量。待测物体如图 6(a)、6(c) 和 6(e) 所示, 通过文中的方法对左右视角采集到的待测物体单

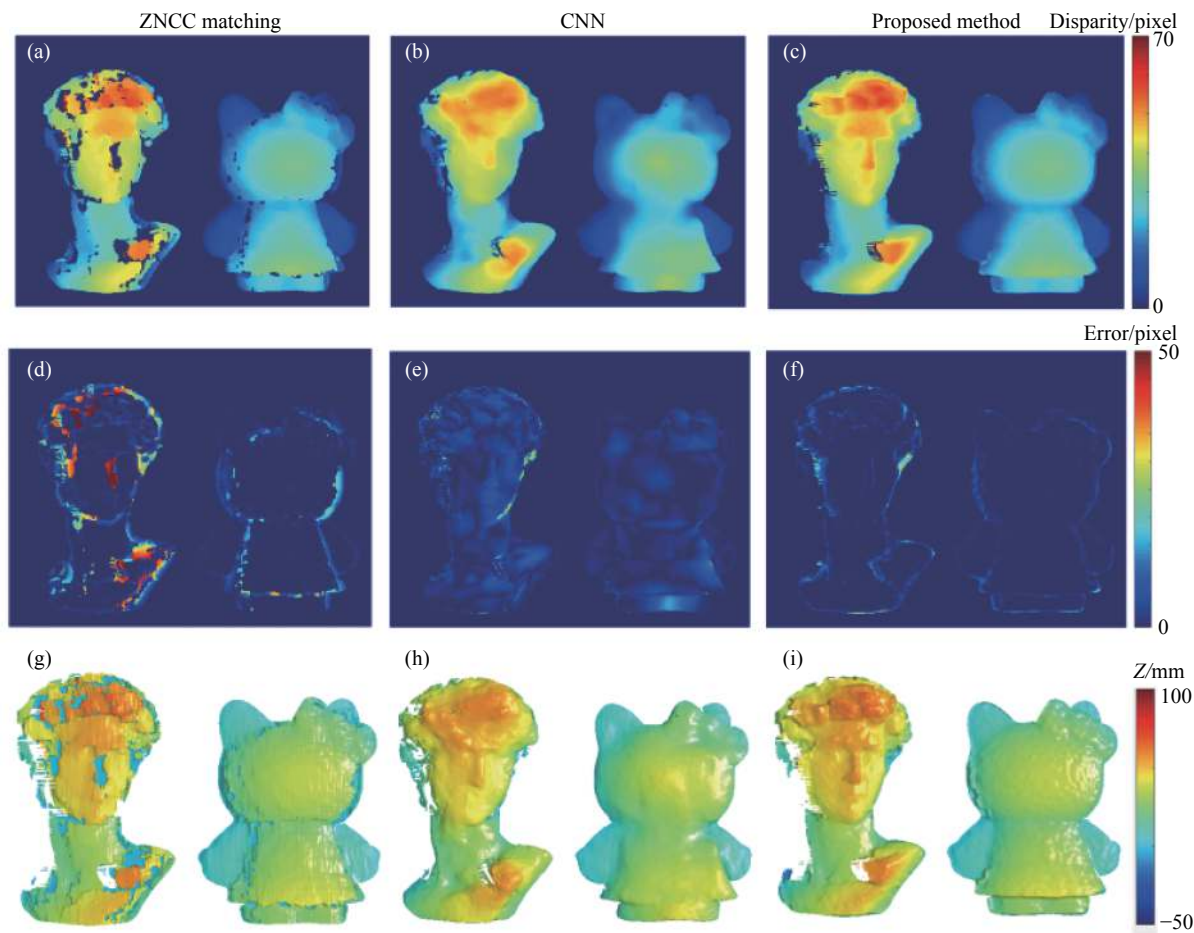


图 5 视差图像和与真实值的误差图。(a) ZNCC 匹配得到的视差图; (b) CNN 网络得到的视差图; (c) 文中方法得到的视差图; (d) ZNCC 匹配的误差图; (e) CNN 网络的误差图; (f) 文中方法的误差图; (g) ZNCC 匹配获得的深度图; (h) CNN 网络获得的深度图; (i) 文中方法获得的深度图

Fig.5 Disparity map and error graph from true value. (a) Disparity of ZNCC matching; (b) disparity of CNN; (c) disparity of the proposed method; (d) error of ZNCC matching; (e) error of CNN; (f) error of the proposed method; (g) depth of ZNCC matching; (h) depth of CNN; (i) depth of the proposed method

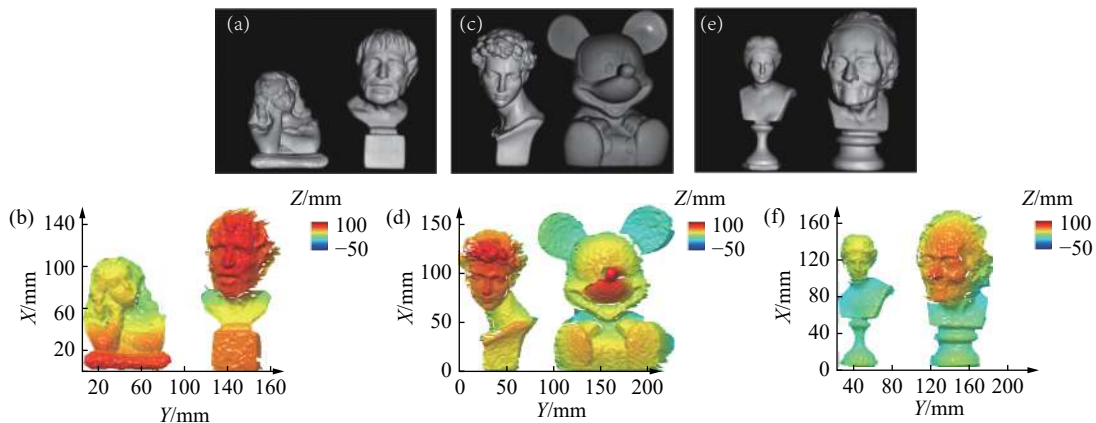


图 6 待测物体与三维重构结果。图 (a)、(c)、(e) 为待测物体; 图 (b)、(d)、(f) 为三维重构结果

Fig.6 Measured object and 3D reconstruction result. (a), (c), (e) measured object; (b), (d), (f) 3D reconstruction result



幅散斑图像进行处理,获得的三维重构结果分别如图 6(b)、6(d) 和 6(f) 所示。从物体的三维重构结果中可以看出,虽然在待测物体边缘位置和细节处的测量精度还有待提高,但是深度学习的方法对六个不同的物体均可以获得其准确的三维数据。经实验结果表明,文中的方法具有较好的鲁棒性和稳定性,对于不同大小和形状的对象都可以获得其准确的三维信息。

为了定量分析基于深度学习的散斑投影轮廓术的三维测量精度。文中对两个标准的陶瓷球进行精度分析。陶瓷球的半径为 25.4 mm,球心距为 100.069 mm。

通过文中的方法获得精度球散斑图像的视差数据,进而获得其三维数据并进行三维重构,两个精度球的三维重构结果如图 7(a) 和图 7(d) 所示。将获得的精度球三维数据与球面拟合的参考数据进行比较,获得陶瓷球测量的误差值,陶瓷球的误差分布如图 7(b) 和图 7(e) 所示,陶瓷球误差直方图分布如图 7(c) 和图 7(f) 所示。经过计算获得的两个陶瓷球均方根误差 (RMS) 分别为: 282.66  $\mu\text{m}$  和 289.58  $\mu\text{m}$ 。通过对标准的陶瓷球的精度分析实验结果可以表明,文中的方法通过仅一张散斑图像就可以实现测量精度约为 290  $\mu\text{m}$  的三维测量。

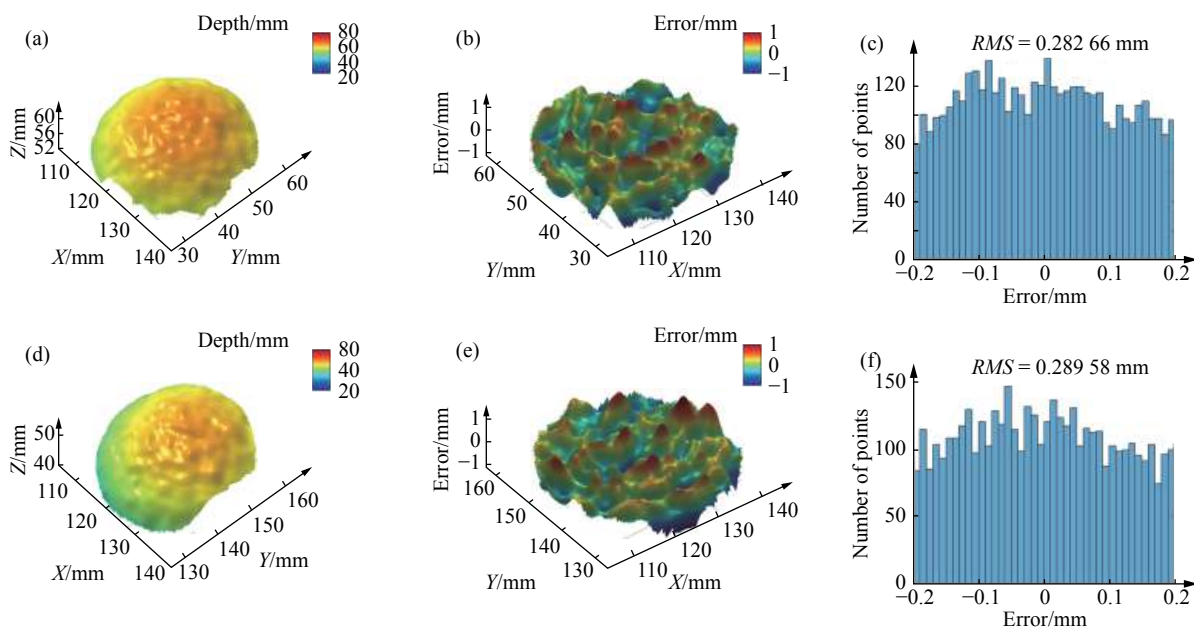


图 7 精度球三维重构及其分析; (a) 图为左边精度球三维重构结果; (b) 图为左边精度球误差分布; (c) 图为左边精度球误差直方图; (d) 图为右边精度球三维重构结果; (e) 图为右边精度球误差分布; (f) 图为右边精度球误差直方图

Fig.7 3D reconstruction and analysis of precision sphere. (a) 3D reconstruction result of left precision sphere; (b) error distribution of left precision sphere; (c) error histogram of left precision sphere; (d) 3D reconstruction result of right precision sphere; (e) error distribution of right precision sphere; (f) error histogram of right precision sphere

## 5 结 论

文中实现了基于深度学习的散斑投影轮廓术。通过对基础的卷积神经网络进行改进,采用了以卷积神经网络为子网络的孪生神经网络结构对左右视图的散斑图像进行匹配,通过将以有效像素点为中心的图像块作为训练数据,并且利用条纹投影轮廓术获得的视差数据对训练数据进行标记的方法完成神经网络的训练和学习过程。利用训练完成的深度学习模

型对单幅散斑图像进行预测,获得散斑图像的匹配聚合代价,并最终获得散斑图像的视差数据和物体的三维数据。为了进一步提高三维测量的精度,利用半全局匹配,左右一致性检验等处理方法对神经网络输出的数据进行后处理。经过实验结果表明,深度学习的方法相比于传统的散斑图像匹配算法可以获得更为准确的物体三维测量结果,并具有较好的鲁棒性和稳定性。此外,通过对精度球进行精度分析,证明文中所提出的方法的三维测量精度约为 290  $\mu\text{m}$ 。从而证

明了,利用深度学习可以仅使用一张散斑图像获得较高精度的三维重构结果。这项技术将有希望用于三维测量精度不高的快速测量领域,如三维人脸识别。文中所使用的深度学习的框架可能还可以适用于其他三维成像技术,例如DIC<sup>[28]</sup>,时空立体匹配<sup>[29]</sup>,运动补偿<sup>[30]</sup>等。当然,文中所实现的散斑投影三维测量在物体边缘和细节处的测量精度还有待提高。如果神经网络结构和深度学习训练方法在未来可以得到进一步优化,使得散斑投影的测量精度可以与条纹投影的测量精度相媲美,最终实现单幅散斑图像的高精度三维测量。

### 参考文献:

- [1] Su X, Zhang Q. Dynamic 3-D shape measurement method: a review [J]. *Optics and Lasers in Engineering*, 2010, 48(2): 191-204.
- [2] Gorthi S S, Rastogi P. Fringe projection techniques: whither we are? [J]. *Optics and Lasers in Engineering*, 2010, 48(2): 133-140.
- [3] Feng S, Zhang L, Zuo C, et al. High dynamic range 3d measurements with fringe projection profilometry: a review [J]. *Measurement Science and Technology*, 2018, 29(12): 122001.
- [4] Zhang S. High-speed 3D shape measurement with structured light methods: A review [J]. *Optics and Lasers in Engineering*, 2018, 106: 119-131.
- [5] Zhou P, Zhu J, Jing H. Optical 3-D surface reconstruction with color binary speckle pattern encoding [J]. *Optics Express*, 2018, 26(3): 3452-3465.
- [6] Guo J, Peng X, Li A, et al. Automatic and rapid whole-body 3D shape measurement based on multinode 3D sensing and speckle projection [J]. *Applied Optics*, 2017, 56(31): 8759-8768.
- [7] Pan B, Xie H, Wang Z. Equivalence of digital image correlation criteria for pattern matching [J]. *Applied Optics*, 2010, 49(28): 5501-5509.
- [8] Liu X, Zhao H, Zhan G, et al. Rapid and automatic 3D body measurement system based on a GPU-Steger line detector [J]. *Applied Optics*, 2016, 55(21): 5539-5547.
- [9] Schaffer M, Grosse M, Harendt B, et al. High-speed three-dimensional shape measurements of objects with laser speckles and acousto-optical deflection [J]. *Optics Letters*, 2011, 36(16): 3097-3099.
- [10] Lionello G, Cristofolini L. A practical approach to optimizing the preparation of speckle patterns for digital-image correlation [J]. *Measurement Science and Technology*, 2014, 25(10): 107001.
- [11] Zhong F, Kumar R, Quan C. RGB laser speckles based 3D profilometry [J]. *Applied Physics Letters*, 2019, 114(20): 201104.
- [12] García J, Zalevsky Z, García-Martínez P, et al. Three-dimensional mapping and range measurement by means of projected speckle patterns [J]. *Applied Optics*, 2008, 47(16): 3032-3040.
- [13] Zhou P, Zhu J, You Z. 3-D face registration solution with speckle encoding based spatial-temporal logical correlation algorithm [J]. *Optics Express*, 2019, 27(15): 21004-21019.
- [14] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
- [15] Rivenson Y, Zhang Y, Günaydin H, et al. Phase recovery and holographic image reconstruction using deep learning in neural networks[J]. *Light: Science & Applications*, 2018, 7(2): 17141.
- [16] Sinha A, Lee J, Li S, et al. Lensless computational imaging through deep learning [J]. *Optica*, 2017, 4(9): 1117-1125.
- [17] Feng S, Chen Q, Gu G, et al. Fringe pattern analysis using deep learning [J]. *Advanced Photonics*, 2019, 1(2): 025001.
- [18] Feng S, Zuo C, Yin W, et al. Micro deep learning profilometry for high-speed 3D surface imaging [J]. *Optics and Lasers in Engineering*, 2019, 121: 416-427.
- [19] Zuo C, Feng S, Huang L, et al. Phase shifting algorithms for fringe projection profilometry: A review [J]. *Optics and Lasers in Engineering*, 2018, 109: 23-59.
- [20] Malacara D. Optical Shop Testing[M]. New York: John Wiley & Sons, 2007, 59.
- [21] Zuo C, Huang L, Zhang M, et al. Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review [J]. *Optics and Lasers in Engineering*, 2016, 85: 84-103.
- [22] Yin W, Chen Q, Feng S, et al. Temporal phase unwrapping using deep learning[J]. arXiv preprint arXiv: 1903.09836, 2019.
- [23] Yin W, Zuo C, Feng S, et al. High-speed 3D shape measurement with the multi-view system using deep learning[C]//SPIE, 2019, 11189: 111890B.
- [24] Zbontar J, LeCun Y, others. Stereo matching by training a convolutional neural network to compare image patches [J]. *Journal of Machine Learning Research*, 2016, 17(1-32): 2.
- [25] Luo W, Schwing A G, Urtasun R. Efficient deep learning for stereo matching[C]//IEEE, 2016: 5695-5703.
- [26] Hirschmuller H. Stereo processing by semiglobal matching and mutual information [J]. *IEEE Transactions on Pattern Analysis*

- and Machine Intelligence*, 2007, 30(2): 328–341.
- [27] Zhang M, Chen Q, Tao T, et al. Robust and efficient multi-frequency temporal phase unwrapping: optimal fringe frequency and pattern sequence selection [J]. *Optics Express*, 2017, 25(17): 20381–20400.
- [28] Pan B. Digital image correlation for surface deformation measurement: historical developments, recent advances and future goals [J]. *Measurement Science and Technology*, 2018, 29(8): 082001.
- [29] Zhang L, Curless B, Seitz S M. Spacetime stereo: Shape recovery for dynamic scenes[C]//IEEE, 2003, 2: II-367.
- [30] Feng S, Zuo C, Tao T, et al. Robust dynamic 3-D measurements with motion-compensated phase-shifting profilometry [J]. *Optics and Lasers in Engineering*, 2018, 103: 127–138.