第 42 卷 第 14 期/2022 年 7 月/光学学报 研究论文

光学学报

基于立体像对数据集的非对称离焦数据集构建

李云鹏[1,2]，葛宝臻[1,2*]，田庆国[1,2]，吕且妮[1,2]

[1]天津大学精密仪器与光电子工程学院，天津 300072；

[2]天津大学光电信息技术教育部重点实验室，天津 300072

摘要 左右图像非对称离焦模糊会导致双目立体视觉系统的立体匹配失败。为训练能够应对图像模糊的神经网络，基于归一化模糊度(NBL)的分层景深叠加算法，以FlyingThings-Stereo立体像对数据集为例，添加随景物深度变化的模糊，构建非对称离焦立体视觉数据集。新建的数据集提供非对称离焦的立体像对，可用于训练去模糊网络或立体匹配网络。在训练去模糊网络时，分别向网络的输入和输出端提供模糊和清晰的立体像对；在训练立体匹配网络时，向网络的输入和输出端提供模糊的立体像对和视差真值。利用虚拟合成和实景拍摄数据对训练后的网络进行验证，结果表明本数据集可以有效训练去模糊和立体匹配神经网络，使其具备应对离焦模糊的能力，实现图像去模糊和基于模糊图像的立体匹配。

关键词 机器视觉；图像处理；非对称离焦；去模糊；数据集；立体匹配

中图分类号 TP391 **文献标志码** A **DOI**：10.3788/AOS202242.1415001

Unbalanced Defocus Dataset Construction Based on Stereo Image Pair Dataset

Li Yunpeng[1,2], Ge Baozhen[1,2*], Tian Qingguo[1,2], Lü Qieni[1,2]

[1]*School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin* 300072, *China*;

[2]*Key Laboratory of Opto-Electronic Information Technology, Ministry of Education, Tianjin University, Tianjin* 300072, *China*

Abstract The unbalanced defocus blur of the left and right images leads to stereo matching failure in a binocular stereo vision system. In order to train a neural network that can deal with the image blur, this paper constructs an unbalanced defocus stereo vision dataset by adding the blur varying with the depth using a normalized blur level (NBL) based layered depth-of-field rendering algorithm and taking the FlyingThings-Stereo image pair dataset as an example. The proposed dataset can provide the unbalanced defocus stereo images and be used to train deblurring or stereo matching networks. When training the deblurring network, the dataset provides blurry and clear stereo images to the network′s input and output ends. When training stereo matching network, fuzzy stereo pairs and parallax truth values are provided to the input and output ends of the network. The network is verified by synthetic and real-scene data after it is trained. Results show that the proposed dataset can effectively train the deblurring and stereo matching neural networks and enables their ability to cope with defocus blur, so as to achieve the image deblurring and stereo matching based on blurry images.

Key words machine vision; image processing; unbalanced defocus; deblurring; dataset; stereo matching

1 引 言

双目立体视觉技术在工业测量[1]、视觉导航[2]、自动驾驶[3]、航空遥感[4]等领域应用广泛。立体匹配作为双目立体视觉技术的核心步骤，其前提条件是位于左、右图像的同名点具有相似的图像特征。然而，左右相机对焦位置不一致时，可能存在非对称离焦模糊[5]，导致立体匹配失败。随着理论和硬件的发展，深度神经网络应用于立体视觉的研究已成为当今的前沿热点[6-10]。数据集对网络的训练至关重要，而已有的立体视觉数据集很少考虑非对称离焦问题。为了使神经网络具备应对图像模糊的能力，数据集中的立体像对需要具备非对称离焦模糊的特点。

目前广泛使用的立体视觉数据集按照构建的方式

收稿日期：2021-12-15；**修回日期**：2022-01-17；**录用日期**：2022-01-20

基金项目：国家自然科学基金(61535008)

通信作者：*gebz@tju.edu.cn

主要分为实景数据集和合成数据集两类。实景数据集一般利用结构光或激光雷达等主动三维测量技术，结合立体相机来同时获取立体图像及其三维坐标。在自动驾驶领域，Geiger等[11]使用高分辨率相机、全角度激光雷达和定位系统组成的车载采集系统，采集了389组立体视觉图像对，建立了KITTI立体视觉数据集和立体匹配算法排行榜；Cordts等[12]采集了更大规模的立体驾驶视频数据，并加入了场景的分类标注。针对室内场景，Scharstein等[13]利用结构光三维技术建立了Middlebury数据集，得到高分辨率的立体图像以及高精度的视差图；Couprie等[14]使用Kinect深度相机在室内拍摄了带有图像与深度信息的视频，并且对景物进行了分类标注。Schöps等[15]针对多视图立体视觉问题，用激光扫描仪、高分辨率单反相机、低分辨率视频录像机采集了一系列室内和户外的图像，建立了ETH3D数据集与匹配算法排行榜。Bao等[16]为改进立体匹配网络在室内场景的效果，建立了含有2050个立体像对的InStereo2K数据集。建立实景数据集耗资巨大，因此数据规模往往受限。Cho等[17]使用手持双目相机拍摄了大量的立体像对，借助立体匹配算法生成视差的伪真值，训练用于单目景深估计的网络。目前，规模有限的实景数据一般仅用于微调预训练的网络；而网络的预训练则通常使用规模更大的合成数据集。

合成数据集使用计算机建模与图形渲染相结合的方法，计算生成虚拟的立体图像和完美的视差图。Bulter等[18]根据开源的三维渲染动画短片建立了Sintel光流数据集和光流算法排行榜。Mayer等[19]为了更好地训练光流和立体匹配网络，建立了具有大规模数据的SceneFlow数据集，它包括三个子数据集：以自由飞行的随机物体组成的FlyingThings3D数据集、以运动的卡通形象为主的Monkaa数据集和仿照驾驶环境的Driving数据集。其中，FlyingThings3D数据集的立体视觉子集FlyingThings3D-Stereo具有数据量大、景物随机性好、视差图准确等优点，被广泛用于立体视觉网络的训练。然而，立体视觉数据集中很少考虑到景深带来的图像模糊问题。针对单目图像中的模糊，D'Andres等[20]使用Lytro光场相机建立了带有22幅离焦模糊图像的实景数据集；Lee等[21]为了训练能够估计模糊核尺寸的神经网络，根据像素深度对图像添加离焦模糊，建立了SYNDOF数据集。这些离焦数据集的规模较小，并未考虑非对称离焦问题，因此建立大规模的离焦模糊数据集具有重要意义。

本文以非对称离焦模糊原理为基础，采用分层叠加景深技术，以FlyingThings3D-Stereo大规模立体视觉数据集为例，添加了非对称离焦模糊，使用归一化模糊度(NBL)控制图像的模糊程度，构建大规模的非对称离焦立体视觉数据集，以训练出能够应对离焦模糊的神经网络。本文代码的网址是：https://gitee.com/psyrocloud/unbalanced_defocus_stereo_dataset。

2　基本原理

2.1　离焦模糊的定量表达

模糊图像I_B可以看作是将清晰图像I与模糊核k进行卷积(⊗)后叠加噪声n形成的[22]：

$$I_B = k \otimes I + n。 \tag{1}$$

不同类型的图像模糊对应不同类型的模糊核k，如运动模糊核一般为不规则的线状，离焦模糊核则呈现二维高斯分布。对于离焦模糊，模糊核半径与景物到对焦平面的距离有关。如图1所示，孔径为D，相机对焦到S点，景物在传感器上成一个锐利点像S'，物距为Z_0，像距为w_0；对焦平面外一点P，将其清晰成像到P'点，对应的物距为Z，像距为w。根据高斯成像公式与相似三角形原理，有：

$$\begin{cases} \dfrac{1}{Z_0}+\dfrac{1}{w_0}=\dfrac{1}{f} \\ \dfrac{1}{Z}+\dfrac{1}{w}=\dfrac{1}{f} \\ \dfrac{2r_B}{\left|w_0-w\right|}=\dfrac{D}{w} \end{cases}, \tag{2}$$

式中：f为镜头焦距；r_B为模糊核半径，可表示为

$$r_B=\left|A\left(\frac{Z_0}{Z}-1\right)\right|, \tag{3}$$

式中：A是系统常数，可表示为

$$A=\frac{fD}{2\left(Z_0-f\right)}, \tag{4}$$

此时式(1)可改写为

$$I_B(u,v)=\left[k\left(r_B\right)\otimes I\right](u,v)+n, \tag{5}$$

式中：$k\left(r_B\right)$是随图像坐标(u,v)变化的模糊核函数，其模糊核半径r_B随(u,v)处景物深度的变化而变化。从式(3)可以看出，当$Z=Z_0$时，模糊核半径$r_B=0$；随着距离Z逐渐远离对焦距离Z_0，模糊核的尺寸逐渐变大。与此同时，对应的像素点处会随着模糊核尺寸的增大而变得更加模糊。

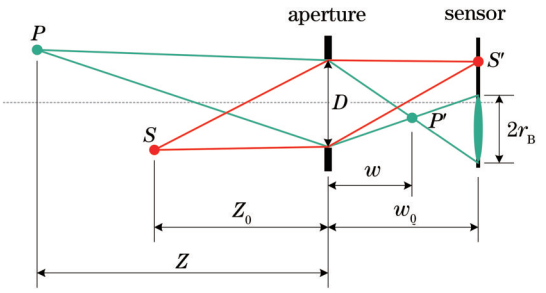

图1　随深度变化的模糊核半径

Fig. 1　Depth-dependent radius of blur kernel

2.2　基于NBL分层叠加景深技术的离焦模糊生成方法

在构建新的数据集之前，首先介绍如何根据视差

向清晰的图像添加离焦模糊。根据式(3)的模型，离焦模糊的程度随景物的深度变化而变化，需要使用分层叠加景深技术[23]添加随深度变化的离焦模糊。该方法的核心思想是：将图像按照对应的景物深度分层，以不同尺寸的模糊核分别对每层图像进行卷积，以添加模糊，最后通过将模糊的分层图像由远及近地叠加，得到完整的离焦模糊图像。

生成模糊核$k(r_B)$的方法有两类：一类从严格的物理模型出发，使用衍射光学理论得到物理精确的点扩展函数[24]；另一类使用几何近似，按照弥散圆的尺寸，将模糊核近似为高斯函数[25]或圆盘函数[26]。本文选择参数量更少的圆盘函数$f_D(x,y)$来近似模糊核：

$$k(r_B)=f_D(x,y)=\begin{cases}\delta(x,y), & \sqrt{x^2+y^2}<r_B\\ 0, & \sqrt{x^2+y^2}\geqslant r_B\end{cases},\tag{6}$$

式中：(x,y)是以模糊核中心为原点的二维坐标；$\delta(x,y)$是冲击函数。圆盘函数在二维平面上的积分是1，即$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f_D(x,y)\mathrm{d}x\mathrm{d}y=1$。

模糊核半径由式(3)决定，受f、D、Z_0、Z等众多系统参数影响。然而，当系统参数未知时，较为合理的选择是让模糊核的半径归一化到$[0,R]$。根据视差公式$Z=fT/d$，$Z_0=fT/d_0$(T为基线宽度，d为视差，d_0为图像对焦平面对应的视差值)，用视差代替深度，将式(3)改写为

$$r(d)=\hat{A}\left(\frac{d}{d_0}-1\right),\tag{7}$$

式中：r为模糊核半径。其中，将最大视差d_X和最小视差d_S分别代入式(7)，结合$R=\max\{r(d_S),r(d_X)\}$，求解得到$\hat{A}$为

$$\hat{A}=\frac{R}{\max\left\{\left|\frac{d_S}{d_0}-1\right|,\left|\frac{d_X}{d_0}-1\right|\right\}},\tag{8}$$

式中：$\hat{A}$在一幅离焦图像中是常数；R是离焦图像中最大的模糊核半径，是控制图像整体模糊程度的关键参数。此处，R是需要人工设置的参数，设置完后才能通过式(8)求解$\hat{A}$。从式(7)中可以发现，$r(d)$的值可以为负，代表此时视差d对应的Z大于Z_0。由于引入了NBL的概念，添加模糊时可以不再考虑系统参数。

分层模糊并叠加时，将像素对应的视差d四舍五入取整为M，其中最大和最小视差是M_X和M_S；按照每一次取整后的视差M对应一个分层的原则，将图像分成M_X-M_S+1层3通道的子图像$I^{(M)}$及其透明度图像$\alpha^{(M)}$，整数$M=\langle d_S\rangle,\langle d_S\rangle+1,\cdots,\langle d_X\rangle$，符号$\langle\bullet\rangle$代表四舍五入取整。当图像$I$中的像素$(u,v)$对应的取整视差等于$M$时，有

$$\begin{cases}I^{(M)}(u,v)=I(u,v)\\ \alpha^{(M)}(u,v)=1\end{cases}。\tag{9}$$

对于其余的像素(u,v)，有

$$\begin{cases}I^{(M)}(u,v)=0\\ \alpha^{(M)}(u,v)=0\end{cases}。\tag{10}$$

根据这些子图像及其透明度图像，计算得到添加离焦模糊的图像B满足

$$\begin{cases}B=B^{(M_X)}/\alpha^{(M_X)}\\ B^{(M+1)}=k(M)\otimes I^{(M)}+\left[1-k(M)\otimes\alpha^{(M)}\right]\odot B^{(M)},\\ B^{(M_S)}=0\end{cases}\tag{11}$$

式中：模糊核$k(M)$是半径为$r(M)$的圆盘函数；$B^{(M)}$是叠加过程中的模糊子图像；符号$\odot$代表像素对像素的点乘操作；$B^{(M_X)}$为添加模糊后的子图像；$\alpha^{(M_X)}$为子图像的透明度图像。

2.3 非对称离焦模糊立体视觉数据集的构建

基于上述方法向图像添加模糊时，仅需指定NBL，即最大模糊核半径R和左、右图像对焦深度对应的视差值d_L、d_R。此时，仅用NBL就可以表征一对图像的模糊度，因为无论左、右图像对焦深度对应的视差值d_L、d_R为多少，算法都会使用式(8)中的$\hat{A}$将模糊核半径归一化到$[0,R]$范围内。

模糊核尺寸的均匀随机分布可让训练出的神经网络具有更好的泛化能力。d_L和d_R虽然不影响图像的模糊程度，但是会决定对焦与离焦景物在图像中的分布。添加模糊时，选择和实际拍摄比较相近的对焦策略，即左、右图像各自对其中某一靠近视野中央的像素(u_L,v_L)、(u_R,v_R)分别对焦，选取其对应的视差作为d_L、d_R。由于FlyingThings-Stereo数据集中的物体是随机分布在空间中的，因此使用固定像素位置得到的对焦视差仍然是随机的。此外，虽然一幅图像中对焦位置是固定的，但是训练神经网络时要在随机位置上截取左、右图像的子区域，因此在训练过程中，左、右图的对焦位置和对焦深度实际上也是随机的，这能够覆盖实际的情况。

原始的FlyingThings-Stereo数据集中包含21818条训练数据和4248条测试数据，每条数据包含左、右立体图像以及左、右视差图；立体图像和视差图的分辨率均为960 pixel×540 pixel。这些图像是通过三维建模、在空间中随机分布并渲染得到的。训练数据和测试数据使用相同的方式生成，场景相似但不重复。

综上，在构建非对称离焦数据集时，对于FlyingThings-Stereo中的所有立体像对，使用相同的NBL（NBL为14）、位于视野中央略微偏左的$(u_L,v_L)=(380,270)$和位于视野中央略微偏右的$(u_R,v_R)=(580,270)$等参数添加非对称离焦模糊，生成新的离焦模糊立体像对。此外，新构建的数据集不改动原有的视差图。

本文构建的数据集在保持FlyingThings-Stereo数据集的结构、分辨率及规模的基础上，为每条数据添加了一对非对称离焦模糊的左、右图像。用于训练的数

据有 21818 条，用于测试的数据有 4248 条。新的数据集中，每条数据包括：1）离焦模糊的左、右图像（NBL为 14，对焦的像素位置固定，对焦的景物深度随机）；2）清晰的左、右图像；3）左、右视差图。以测试集中第 0 条和第 2000 条数据为例，图 2 给出了构建的数据集的数据样例。其中，图 2（a1）、（a2）、（b1）、（b2）中下方的虚线框和实线框内的图像是模糊图像中对应区域的放大图，展示了非对称离焦的情况。

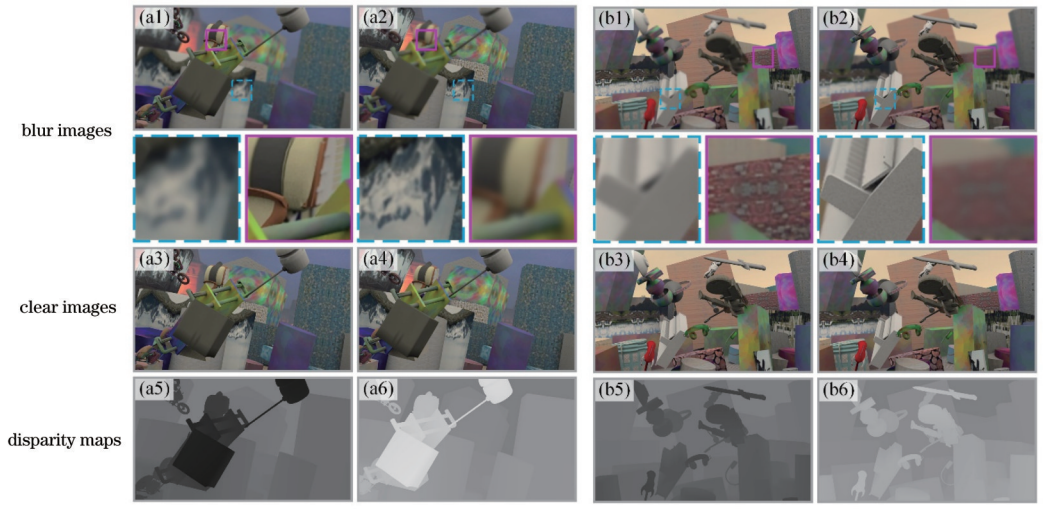

图 2 非对称离焦模糊立体视觉数据集的数据样例。（a1）～（a6）第 0 条数据模糊的左、右图像，清晰的左、右图像，以及左、右视差图；（b1）～（b6）第 2000 条数据模糊的左、右图像，清晰的左、右图像，以及左、右视差图

Fig. 2 Data samples in unbalanced defocus stereo vision dataset. (a1)-(a6) Left and right blur images, left and right clear images, and left and right disparity maps of No. 0 data; (b1)-(b6) left and right blur images, left and right clear images, and left and right disparity maps of No. 2000 data

3 实验与分析

为验证构建数据集的有效性，使用本数据集训练了多种神经网络，并用测试数据集和实景数据分别验证训练出的网络是否有效以及是否具有泛化能力。

3.1 基于去模糊网络的训练实验

本节验证所构建的数据集训练去模糊神经网络的有效性。使用本文所构建的数据集，分别训练了立体去离焦模糊的网络 BLNet[5]、单目去运动模糊的网络[6]（简称 Nah），以及立体去运动模糊的网络 DavaNet[7]。训练的硬件是两块 NVidia GTX 2080TI 图形计算卡，训练过程中将图像随机裁剪为 256 pixel×256 pixel 分辨率的子图像，并添加均值为 0、标准差为 0.01、动态范围是[0, 1]的随机高斯噪声，训练的批大小是 10，优化求解器是 Adam，其中，一阶矩估计的指数衰减率 $\beta_1=0.9$，二阶矩估计的指数衰减率 $\beta_2=0.999$，训练代数是 250 代，损失函数分别与文献[5]、文献[6]、文献[7]保持一致。在评价图像去模糊效果时，使用峰值信噪比（PSNR）和结构相似性（SSIM）这两个指标，它们的值越高，说明去模糊的效果越好。对于每组立体像对，选择左、右图像的平均 PSNR 和 SSIM 值作为评价指标。

用于验证去模糊效果的合成图像是从非对称离焦数据集的验证集前 4000 条数据中每次间隔 500 条数据抽取的，共包含 9 条数据。合成图像的去模糊结果如表 1 所示，从 PSNR 和 SSIM 的均值来看，所选的三种去模糊网络均成功地学会了去除立体图像的非对称离焦模糊。从 PSNR 来看，Nah 的单目方法最为有效。但是从图 3 中局部放大的可视化结果来看，单目去模糊无法保证左、右图像之间去模糊效果的一致性。而 DavaNet 和 BLNet 等立体去模糊网络则可以更好地在去模糊的同时，保持左、右图像清晰度的一致性。其中 BLNet 保持清晰度一致的视觉效果更好。从 SSIM 来看，Nah 和 DavaNet 更具优势。综上，从表 1 和图 3 可以看出，本文数据集有效地训练了去模糊网络。

表 1 合成图像的去模糊结果

Table 1 Deblurred results of synthetic image

Number	Nah		DavaNet		BLNet	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
0	36.30	0.94	35.50	0.93	33.49	0.93
500	39.20	0.97	37.87	0.96	35.38	0.96
1000	37.22	0.97	36.74	0.96	33.58	0.95
1500	34.47	0.93	33.42	0.92	32.29	0.91
2000	36.02	0.95	34.97	0.94	32.44	0.93
2500	35.00	0.94	33.99	0.92	31.46	0.91
3000	33.54	0.92	32.80	0.90	31.95	0.90
3500	33.52	0.93	32.71	0.92	31.65	0.92
4000	34.32	0.96	33.93	0.95	32.47	0.94
Average	35.51	0.95	34.66	0.93	32.75	0.93

为了进一步验证所提数据集的泛化性能，在 Middlebury 2014 实景立体视觉数据集中加入了非对称离焦模糊，将所有左图像对焦在远景处、右图像对焦在近景处，以测试本文数据集训练得到的网络是否具

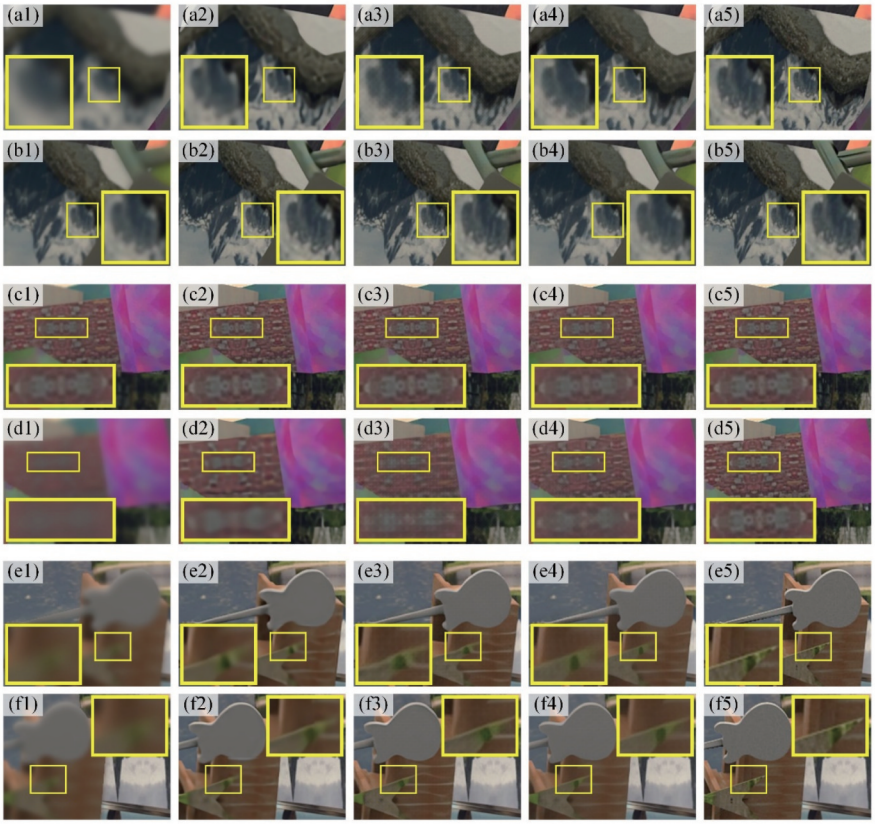

图 3 合成图像去模糊的可视化结果。(a1)～(a5)第 0 条数据模糊的左图像，Nah、DavaNet、BLNet 去模糊后的左图像，以及清晰的左图像；(b1)～(b5)第 0 条数据模糊的右图像，Nah、DavaNet、BLNet 去模糊后的右图像，以及清晰的右图像；(c1)～(c5)第 2000 条数据模糊的左图像，Nah、DavaNet、BLNet 去模糊后的左图像，以及清晰的左图像；(d1)～(d5)第 2000 条数据模糊的右图像，Nah、DavaNet、BLNet 去模糊后的右图像，以及清晰的右图像；(e1)～(e5)第 4000 条数据模糊的左图像，Nah、DavaNet、BLNet 去模糊后的左图像，以及清晰的左图像；(f1)～(f5)第 4000 条数据模糊的右图像，Nah、DavaNet、BLNet 去模糊后的右图像，以及清晰的右图像

Fig. 3 Visualized deblurred results of synthetic image. (a1)-(a5) Left blur image, deblurred left images by Nah, DavaNet, and BLNet, and clear left image of No. 0 data; (b1)-(b5) right blur image, deblurred right images by Nah, DavaNet, and BLNet, and clear right image of No. 0 data; (c1)-(c5) left blur image, deblurred left images by Nah, DavaNet, and BLNet, and clear left image of No. 2000 data; (d1)-(d5) right blur image, deblurred right images by Nah, DavaNet, and BLNet, and clear right image of No. 2000 data; (e1)-(e5) left blur image, deblurred left images by Nah, DavaNet, and BLNet, and clear left image of No. 4000 data; (f1)-(f5) right blur image, deblurred right images by Nah, DavaNet, and BLNet, and clear right image of No. 4000 data

有泛化性能。在排除了视差范围过大的两幅图像后，使用数据集中其余所有 8 幅图像进行测试，其结果如表 2 所示，可以看到表 2 中的 PSNR 和 SSIM 值的高低趋势与表 1 所示一致。部分可视化结果如图 4 所示，从局部放大的方框中可以看出 3 种网络均有效地去除了非对称离焦模糊，说明本文构建的数据集可训练出具备泛化性能的去模糊网络。

3.2 基于 PSMNet 立体匹配网络的训练实验

为了验证数据集训练立体匹配网络的有效性，使用本文提出的模糊数据集训练经典的立体匹配网络 PSMNet[8]，将得到的网络称为 PSMNet-B。同时，使用清晰数据集训练得到的 PSMNet-C 网络与 PSMNet-B 网络进行对比，以验证数据集的有效性。使用两块 NVidia RTX 2080TI 图形计算卡进行训练，将图像随机裁剪为 512 pixel×512 pixel 分辨率的子图像，训练的批大小是 4，优化求解器是 Adam，一阶矩估

表 2 Middlebury 2014 数据集实景图像的去模糊结果

Table 2 Deblurred results of real-scene images in Middlebury 2014 dataset

Scene	Nah		DavaNet		BLNet	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Adirondack	38.28	0.96	36.27	0.94	32.68	0.94
Motorcycle	30.01	0.92	29.55	0.91	27.47	0.90
Piano	35.36	0.95	33.82	0.92	31.93	0.92
Pipes	31.79	0.90	31.55	0.89	30.00	0.89
Playroom	32.00	0.94	31.04	0.91	30.55	0.91
Playtable	33.31	0.87	30.97	0.83	29.35	0.82
Recycle	37.54	0.96	36.10	0.94	31.00	0.95
Shelves	33.45	0.94	32.88	0.92	29.41	0.92
Average	33.97	0.93	32.77	0.91	30.30	0.90

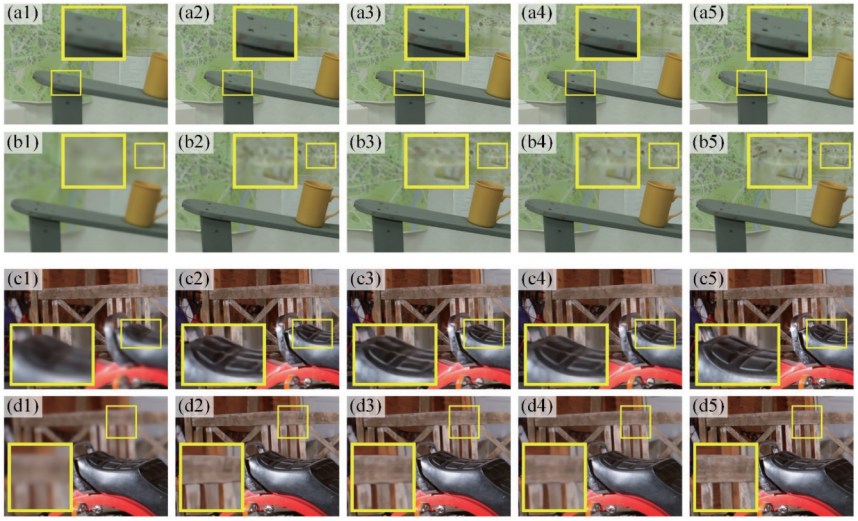

图 4　Middlebury 2014 数据集实景图像去模糊的可视化结果。(a1)～(a5)实景 Adirondack 模糊的左图像，Nah、DavaNet、BLNet 去模糊后的左图像，以及清晰的左图像；(b1)～(b5)实景 Adirondack 模糊的右图像，Nah、DavaNet、BLNet 去模糊后的右图像，以及清晰的右图像；(c1)～(c5)实景 Motorcycle 模糊的左图像，Nah、DavaNet、BLNet 去模糊后的左图像，以及清晰的左图像；(d1)～(d5)实景 Motorcycle 模糊的右图像，Nah、DavaNet、BLNet 去模糊后的右图像，以及清晰的右图像

Fig. 4　Visualized deblurred results of real-scene images in Middlebury 2014 dataset. (a1)-(a5) Left blur image, deblurred left images by Nah, DavaNet, and BLNet, and clear left image of Adirondack; (b1)-(b5) right blur image, deblurred right images by Nah, DavaNet, and BLNet, and clear right image of Adirondack; (c1)-(c5) left blur image, deblurred left images by Nah, DavaNet, and BLNet, and clear left image of Motorcycle; (d1)-(d5) right blur image, deblurred right images by Nah, DavaNet, and BLNet, and clear right image of Motorcycle

计的指数衰减率 $\beta_1=0.9$，二阶矩估计的指数衰减率 $\beta_2=0.999$，训练代数是 10，损失函数使用柔性一范数。评价立体匹配结果时，使用终点误差(EPE，单位是 pixel)和三像素误差(D3，即 EPE 大于 3 的像素占图像总像素的比例，以百分数表示)这两个指标，它们的值越低，表示匹配结果越好。

使用非对称离焦数据集的测试集中的部分图像进行测试，匹配结果如表 3 所示，可以发现：直接用 PSMNet-C 匹配的结果中，平均 D3 达到 34.49%；而使用 PSMNet-B 网络得到的平均 D3 为 9.16%，这相较于 PSMNet-C 的 D3，正确率提高了约 73.4%。与此同时，EPE 也从 6.96 pixel 下降到 2.01 pixel，提升了约 71.1%。这说明 PSMNet-B 的结果明显优于 PSMNet-C。将表 3 中的部分匹配结果进行可视化，如图 5 所示，视差图中的纯黑和纯白像素分别表示无效的遮挡区域和匹配错误的区域，纯白区域的面积越小，匹配错误率越低；其视觉感受与表 3 数据一致，在视差图中，方框部分的差异尤为明显。实验说明本文提出的数据集可以让 PSMNet 在带有模糊的图像上进行匹配，与使用清晰图像训练的结果相比，本文方法的效果大幅提升。

为进一步验证本文数据集能否使得训练的网络具有泛化性，使用添加了非对称离焦的 Middlebury 数据集进行实景图像立体匹配测试。测试结果如表 4 所示，从 D3 和 EPE 来看，PSMNet-B 相比于 PSMNet-C

表 3　合成图像的立体匹配结果

Table 3　Stereo matching results of the synthetic data

Number	PSMNet-C		PSMNet-B	
	D3 /%	EPE /pixel	D3 /%	EPE /pixel
0	40.90	9.25	13.20	2.63
500	21.30	3.16	2.20	0.74
1000	14.90	5.63	10.30	3.62
1500	42.60	9.71	13.50	2.53
2000	46.50	11.61	10.20	1.71
2500	29.30	5.53	7.20	1.39
3000	37.90	4.82	13.10	2.73
3500	37.30	6.77	8.30	1.81
4000	39.70	6.20	4.40	0.90
Average	34.49	6.96	9.16	2.01

表 4　Middlebury 2014 数据集实景图像的立体匹配结果

Table 4　Stereo matching results of real-scene images from Middlebury 2014 dataset

Scene	PSMNet-C		PSMNet-B	
	D3 /%	EPE /pixel	D3 /%	EPE /pixel
Adirondack	42.50	7.11	15.20	2.97
Motorcycle	40.10	7.06	20.70	4.42
Piano	33.40	3.89	27.20	5.45
Pipes	57.20	13.56	25.60	6.95
Playroom	44.60	7.92	39.30	9.91
Playtable	37.00	5.16	25.30	5.87
Recycle	31.90	3.96	17.80	2.77
Shelves	63.10	9.01	40.70	6.81
Average	43.73	7.21	26.48	5.64

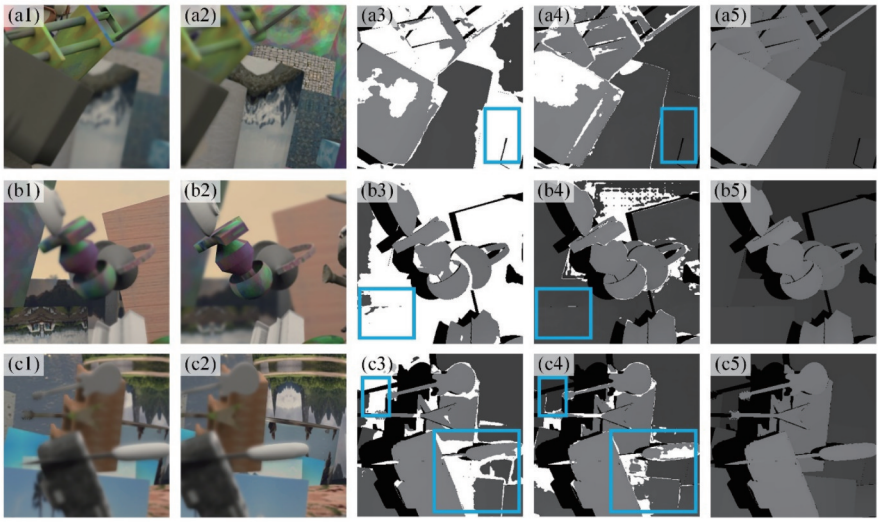

图 5 合成图像的可视化立体匹配结果。(a1)～(a5)第 0 条数据模糊的左、右图像，PSMNet-C、PSMNet-B 得到的视差图，以及视差图的真值；(b1)～(b5)第 2000 条数据模糊的左、右图像，PSMNet-C、PSMNet-B 得到的视差图，以及视差图的真值；(c1)～(c5)第 4000 条数据模糊的左、右图像，PSMNet-C、PSMNet-B 得到的视差图，以及视差图的真值

Fig. 5 Visualized stereo matching results of synthetic image. (a1)-(a5) Left and right blur images, disparity maps of PSMNet-C and PSMNet-B, and ground-truth disparity map of No. 0 data; (b1)-(b5) left and right blur images, disparity maps of PSMNet-C and PSMNet-B, and ground-truth disparity map of No. 2000 data; (c1)-(c5) left and right blur images, disparity maps of PSMNet-C and PSMNet-B, and ground-truth disparity map of No. 4000 data

仍然具有明显的优势，可以认为本文的数据集可以让训练的立体匹配网络具备泛化性。部分结果的可视化效果如图 6 所示，视差图中的纯黑和纯白像素分别表示无效的遮挡区域和匹配错误的区域，纯白区域的面积越小，匹配错误率越低；其视觉感受与表 4 数据保持一致，方框部分的结果差异显著。实验结果表明利用本文的数据集可以训练得到具有泛化性能的立体匹配网络。

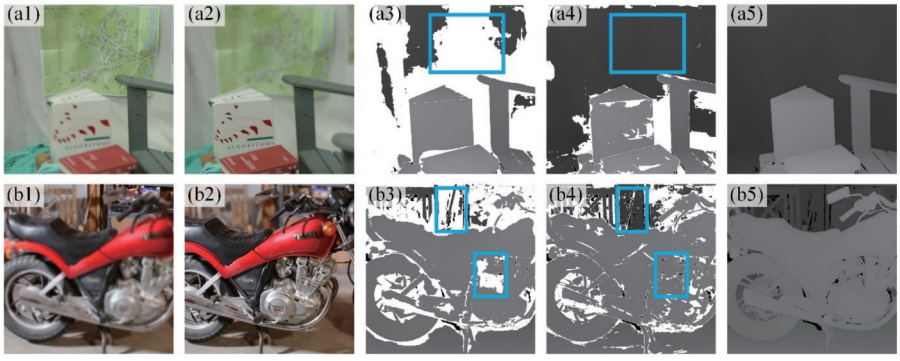

图 6 Middlebury 2014 数据集实景图像的可视化立体匹配结果。(a1)～(a5)实景 Adirondack 的模糊的左、右图像，采用 PSMNet-C、PSMNet-B 得到的视差图，以及视差图的真值；(b1)～(b5)实景 Motorcycle 的模糊的左、右图像，采用 PSMNet-C、PSMNet-B 得到的视差图，以及视差图的真值

Fig. 6 Visualized stereo matching results of real-scene images in Middlebury 2014 dataset. (a1)-(a5) Left and right blur images, disparity maps of PSMNet-C and PSMNet-B, and ground-truth disparity map of Adirondack; (b1)-(b5) left and right blur images, disparity maps of PSMNet-C and PSMNet-B, and ground-truth disparity map of Motorcycle

3.3 实际离焦模糊图的测试实验

为了验证使用新建数据集对实际离焦模糊图的处理效果，使用 3.1 节训练好的去模糊效果较好的 BLNet 和 PSMNet-B，分别对实际拍摄的离焦模糊图像进行去模糊和立体匹配实验。图 7(a)所示为实际的立体视觉相机，相机型号是 Sony A7R2，图像分辨率为 3976 pixel×2652 pixel，镜头焦距为 70 mm，光圈数为 $F/4.0$，左、右相机的基线距离为 160 mm，拍摄距离为 2 m。图 7(b)所示为用于测试的实际场景，是放置在不同深度位置的两个彩色纸盒。拍摄时，左、右相机分别对焦在左、右纸盒上；图 7(c)、(d)是裁剪掉边缘背景的模糊图像，分辨率为 2125 pixel×1144 pixel，图像下方是对应的局部放大图，显示出左、右图像对焦在不同位置而产生的离焦模糊。图 7(e)、(f)是采用 BLNet 去模糊后的结果，对比去模糊前后左图的实线框和右图的虚线框，可以看出网络有效地去除了离焦模糊。图 7(g)是使用 PSMNet-B 处理得到的视差图，图 7(h)为重建的彩色三维点云图，可以看出立体匹配的有效性。

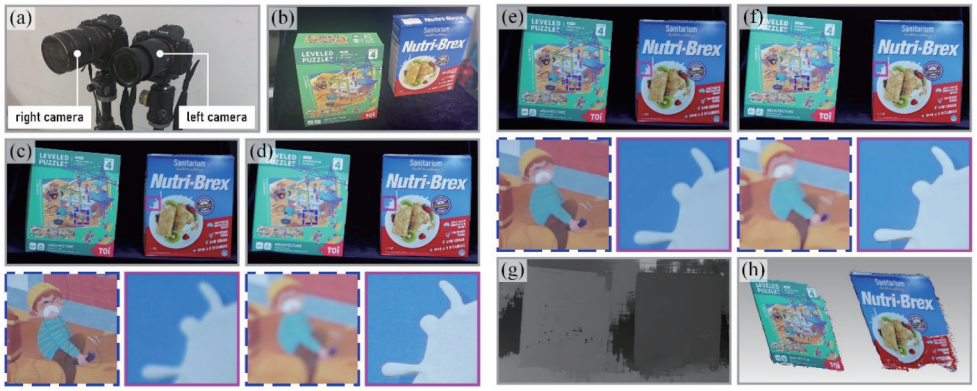

图 7 实际离焦模糊图的测试实验。(a)立体视觉相机;(b)用于测试的实际场景;(c)(d)左、右模糊图像;(e)(f)使用 BLNet 去模糊后的左、右图像;(g)使用 PSMNet-B 匹配得到的视差图;(h)重建的三维点云图

Fig. 7 Test on real defocus blur images. (a) Stereo vision cameras; (b) experimental scene for test; (c)(d) left and right blur images; (e)(f) deblurred left and right images by BLNet; (g) disparity map calculated by PSMNet-B; (h) reconstructed 3D point clouds

4 结 论

针对立体视觉中的非对称离焦问题,提出一种基于 FlyingThings-Stereo 立体视觉数据集的非对称离焦模糊立体数据集的构建方法。使用构建的数据集训练了单目和立体去模糊网络,使其具备图像去模糊的能力;训练了立体匹配神经网络,使其可以容忍非对称离焦模糊,提升了存在图像模糊时立体匹配的精度。通过 Middlebury 2014 数据集实景图像和实际离焦模糊图的测试实验,验证了新建数据集对去模糊网络和立体匹配网络训练的有效性,该数据集可以为后续的去模糊、立体匹配等网络的研究提供有力的补充和支撑。

参 考 文 献

[1] 李承杭, 薛俊鹏, 郎威, 等. 基于相位映射的双目视觉缺失点云插补方法[J]. 光学学报, 2020, 40(1): 0111019.
Li C H, Xue J P, Lang W, et al. Method for interpolation of missing point cloud based on phase mapping in binocular vision[J]. Acta Optica Sinica, 2020, 40(1): 0111019.

[2] 林志林, 张国良, 姚二亮, 等. 动态场景下基于运动物体检测的立体视觉里程计[J]. 光学学报, 2017, 37(11): 1115001.
Lin Z L, Zhang G L, Yao E L, et al. Stereo visual odometry based on motion object detection in the dynamic scene[J]. Acta Optica Sinica, 2017, 37(11): 1115001.

[3] 田苗, 关棒磊, 孙放, 等. 一种无公共视场的多相机系统相对位姿解耦估计方法[J]. 光学学报, 2021, 41(5): 0515001.
Tian M, Guan B L, Sun F, et al. Decoupling relative pose estimation method for non-overlapping multi-camera system[J]. Acta Optica Sinica, 2021, 41(5): 0515001.

[4] 李傲, 唐新明, 祝小勇. 基于统一验证场法的国产高分辨率卫星影像几何定位精度评估[J]. 光学学报, 2021, 41(3): 0328001.
Li A, Tang X M, Zhu X Y. Geometric positioning accuracy evaluation of domestic high-resolution satellite images based on unified verification field method[J]. Acta Optica Sinica, 2021, 41(3): 0328001.

[5] Li Y P, Ge B Z, Tian Q G, et al. Eliminating unbalanced defocus blur with a binocular linkage network [J]. Applied Optics, 2021, 60(5): 1171-1181.

[6] Nah S, Kim T H, Lee K M. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 257-265.

[7] Zhou S C, Zhang J W, Zuo W M, et al. DAVANet: stereo deblurring with view aggregation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 10988-10997.

[8] Chang J R, Chen Y S. Pyramid stereo matching network [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5410-5418.

[9] 陈其博, 葛宝臻, 李云鹏, 等. 基于多重注意力机制的立体匹配算法[J]. 激光与光电子学进展, 2022, 59(16): 1633001.
Chen Q B, Ge B Z, Li Y P, et al. Stereo matching algorithm based on multi attention mechanism[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1633001.

[10] 程鸣洋, 盖绍彦, 达飞鹏. 基于注意力机制的立体匹配网络研究[J]. 光学学报, 2020, 40(14): 1415001.
Cheng M Y, Gai S Y, Da F P. A stereo-matching neural network based on attention mechanism[J]. Acta Optica Sinica, 2020, 40(14): 1415001.

[11] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.

[12] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 3213-3223.

[13] Scharstein D, Hirschmüller H, Kitajima Y, et al. High-resolution stereo datasets with subpixel-accurate ground truth[M]//Jiang X Y, Hornegger J, Koch R. Pattern recognition. Lecture notes in computer science. Cham: Springer, 2014, 8753: 31-42.

[14] Couprie C, Farabet C, Najman L, et al. Indoor semantic segmentation using depth information[EB/OL]. (2013-03-14)[2021-12-13]. https://arxiv.org/abs/1301.3572.

[15] Schöps T, Schönberger J L, Galliani S, et al. A multi-view stereo benchmark with high-resolution images and multi-camera videos[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2538-2547.

[16] Bao W, Wang W, Xu Y H, et al. InStereo2K: a large real dataset for stereo matching in indoor scenes[J]. Science China Information Sciences, 2020, 63(11): 212101.

[17] Cho J, Min D B, Kim Y, et al. Deep monocular depth estimation leveraging a large-scale outdoor stereo dataset [J]. Expert Systems with Applications, 2021, 178: 114877.

[18] Butler D J, Wulff J, Stanley G B, et al. A naturalistic open source movie for optical flow evaluation[M]//Fitzgibbon A, Lazebnik S, Perona P, et al. Computer vision-ECCV 2012. Lecture notes in computer science. Heidelberg: Springer, 2012, 7577: 611-625.

[19] Mayer N, Ilg E, Häusser P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4040-4048.

[20] D'Andres L, Salvador J, Kochale A, et al. Non-parametric blur map regression for depth of field extension[J]. IEEE Transactions on Image Processing, 2016, 25(4): 1660-1673.

[21] Lee J Y, Lee S, Cho S, et al. Deep defocus map estimation using domain adaptation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 12214-12222.

[22] 李海波, 邵文泽. 图像盲去模糊综述: 从变分方法到深度模型以及延伸讨论[J]. 南京邮电大学学报(自然科学版), 2020, 40(5): 84-94.
Li H B, Shao W Z. Blind image deblurring: an overview from variational approaches to deep representation models and beyond[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2020, 40(5): 84-94.

[23] Lee S, Kim G J, Choi S. Real-time depth-of-field rendering using point splatting on per-pixel layers[J]. Computer Graphics Forum, 2008, 27(7): 1955-1962.

[24] Wu Y C, Boominathan V, Chen H J, et al. PhaseCam 3D: earning phase masks for passive single view depth estimation[C]//2019 IEEE International Conference on Computational Photography, May 15-17, 2019, Tokyo, Japan. New York: IEEE Press, 2019: 18793739.

[25] Li F, Sun J, Wang J, et al. Dual-focus stereo imaging [J]. Journal of Electronic Imaging, 2010, 19(4): 043009.

[26] McGuire M, Matusik W, Pfister H, et al. Defocus video matting[J]. ACM Transactions on Graphics, 2005, 24 (3): 567-576.