计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(7): 1771−1785, 2025

多尺度代价聚合的多聚焦图像3维形貌重建框架

闫 涛[1,2] 尚起慧[1,2] 吴 鹏[1] 张江峰[1,2] 钱宇华[1] 陈 斌[3,4]

[1]（山西大学大数据科学与产业研究院 太原 030006）

[2]（山西大学计算机与信息技术学院 太原 030006）

[3]（哈尔滨工业大学重庆研究院 重庆 401151）

[4]（哈尔滨工业大学（深圳）国际人工智能学院 广东深圳 518055）

（hongyanyutian@sxu.edu.cn）

Multi-Scale Cost Aggregation Framework for 3D Shape Reconstruction from Multi-Focus Images

Yan Tao[1,2], Shang Qihui[1,2], Wu Peng[1], Zhang Jiangfeng[1,2], Qian Yuhua[1], and Chen Bin[3,4]

[1]（*Institute of Big Data Science and Industry*, *Shanxi University*, *Taiyuan* 030006）

[2]（*School of Computer and Information Technology*, *Shanxi University*, *Taiyuan* 030006）

[3]（*Chongqing Research Institute of Harbin Institute of Technology*, *Chongqing* 401151）

[4]（*International Research Institute for Artificial Intelligence*, *Harbin Institute of Technology* (*Shenzhen*), *Shenzhen*, *Guangdong* 518055）

Abstract 3D shape reconstruction aims to recover the 3D structure information of the scene by using image sequences with different focus levels. Most of the existing 3D shape reconstruction methods evaluate the focus level of the image sequence from a single scale, and guide the reconstruction process by introducing regularization or post-processing methods. Due to the limitation of the selection space of depth information, the reconstruction results often cannot converge effectively. To address this issue, we propose a multi-scale cost aggregation framework for shape from focus (MSCAS). Firstly, non-downsampling multi-scale transformation is introduced to increase the depth information selection space of the input image sequence, and then the cost aggregation is performed by combining the intra-scale sequence correlation and the inter-scale information constraint. Through this expansion-aggregation mode, the doubling of scene depth representation information and the effective fusion of cross-scale and cross-sequence representation information are realized. As a general framework, MSCAS framework can embed existing model design methods and deep learning methods to achieve performance improvement. The experimental results show that MSCAS framework in this paper reduces the root mean square error (*RMSE*) on average by 14.91% and improves the structural similarity (*SSIM*) by 56.69% in the four datasets after embedding the model design class SFF method. After embedding the deep learning class SFF method, *RMSE* in the four datasets decreases by an average of 1.55% and *SSIM* increases by an average of 1.61%. These results verify the effectiveness of MSCAS framework.

收稿日期：2023-12-05；**修回日期**：2025-01-02

基金项目：国家自然科学基金重点项目（62136005）；国家自然科学基金项目（62472268）；科技创新2030-重大项目（2021ZD0112400）；中央引导地方科技发展资金项目（YDZJSX20231C001, YDZJSX20231B001）；山西省省筹资金资助回国留学人员科研项目（2024-020）；山西省研究生教育创新计划项目（2024KY036）

This work was supported by the Key Program of the National Key Natural Science Foundation of China (62136005), the National Natural Science Foundation of China (62472268), the National Key Research and Development Program of China (2021ZD0112400), the Funds for Central Government Guided Local Science and Technology Development of China (YDZJSX20231C001, YDZJSX20231B001), the Research Project Supported by Shanxi Scholarship Council of China (2024-020), and the Graduate Education Innovation Project of Shanxi Province (2024KY036).

通信作者：钱宇华（jinchengqyh@sxu.edu.cn）

Key words multi-scale cost aggregation; depth estimation; focus measure; multi-focus image sequence; 3D shape reconstruction

摘 要 多聚焦图像3维形貌重建旨在利用不同聚焦水平的图像序列恢复场景的3维结构信息.现有的3维形貌重建方法大多从单一尺度对图像序列的聚焦水平进行评价,通过引入正则化或后处理方法引导重建过程,由于深度信息选择空间的局限性往往导致重建结果无法有效收敛.针对上述问题,提出一种多尺度代价聚合的多聚焦图像3维形貌重建框架(multi-scale cost aggregation framework for 3D shape reconstruction from multi-focus images, MSCAS),该框架首先引入非降采样的多尺度变换增加输入图像序列的深度信息选择空间,然后联合尺度内序列关联与尺度间信息约束进行代价聚合,通过这种扩张-聚合模式实现了场景深度表征信息的倍增与跨尺度和跨序列表征信息的有效融合.作为一种通用框架,MSCAS框架可实现已有模型设计类方法和深度学习类方法的嵌入进而实现性能提升.实验结果表明:MSCAS框架在嵌入模型设计类SFF方法后4组数据集中的均方根误差(root mean squared error, *RMSE*)平均下降14.91个百分点,结构相似度(structural similarity index measure, *SSIM*)平均提升56.69个百分点,嵌入深度学习类SFF方法后4组数据集中的*RMSE*平均下降1.55个百分点,*SSIM*平均提升1.61个百分点.验证了MSCAS框架的有效性和通用性.

关键词 多尺度代价聚合;深度估计;聚焦测量;多聚焦图像序列;3维形貌重建

中图法分类号 TP391.41

DOI: 10.7544/issn1000-1239.202330984 **CSTR:** 32373.14.issn1000-1239.202330984

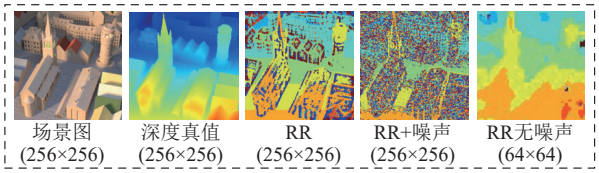
深度估计是计算机视觉和3维形貌重建领域的重要任务之一,目前基于图像的深度估计主要包括单目与多目深度估计.其中,单目深度估计主要包括基于聚焦信息深度估计(shape from focus, SFF)[1],还有基于阴影[2]、纹理[3]和虚焦[4]等线索的深度估计方法;多目深度估计主要包括双目视觉[5]和4D光场[6]等深度估计方法.由于双目视觉深度估计的准确性依赖于配准算法的有效性,而4D光场方法则依赖于复杂的硬件设备,因此多目深度估计通常缺乏场景普适性.单目深度估计方法因其数据采集端无差别化与数据易获取等特点,近年来广受学术界与工业界的关注[7].鉴于图像的阴影、纹理和轮廓等深度线索大多采用单幅图像估计场景深度信息,较少的先验信息使其无法适用于高精度的3维形貌重建场景.而SFF重建则采用多幅不同聚焦水平的图像序列进行场景深度估计,可实现诸多场景深度更为精确的深度信息估计,广泛应用于精密制造产品质量控制、大气环境颗粒物形态学分析、生物医学分子结构观测等领域[8].

SFF重建首先通过等间隔调整相机与重建场景之间的距离,获得可以覆盖重建场景全部景深范围的不同聚焦水平的图像序列,然后采用聚焦测量方法计算图像序列的聚焦水平,最后利用后处理方法融合所有位置最清晰像素的索引,进而得到场景的3维形貌重建结果.然而现有的聚焦测量方法具有典型的场景偏好性,使其无法有效应对复杂场景中的噪声和稀疏纹理的干扰,导致深度信息估计容易出现偏差;与此同时,现有的后处理方法难以有效联合图像序列中的关联关系与不同尺度间互补信息,无法捕捉不同场景的多尺度特征,导致深度误差被放大和积累,降低了场景的重建精度[9].

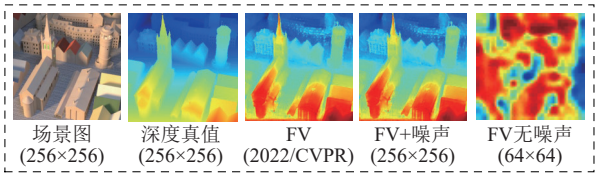
图1展示了不同尺度和噪声对场景深度估计结果造成的影响.图1(a)为模型设计类RR算法[10](2021/TIP)在加入高斯噪声与尺度缩小为64×64时的重建结果;图1(b)为深度学习类FV[11]方法(2022/CVPR)在加入高斯噪声和尺度缩小为64×64时的重建结果.由图1的结果可知,相比深度学习的FV方法,模型设计类RR算法在降低尺度后重建结果有一定优势,但无法有效克服噪声的干扰,而FV方法则有一定的噪声鲁棒性,但具有典型的尺度敏感性.

综上所述,为应对现有SFF方法出现的噪声和尺度鲁棒性等问题,本文提出了一种基于多尺度代价聚合的多聚焦图像3维形貌重建框架(multi-scale cost aggregation framework for 3D shape reconstruction from multi-focus images, MSCAS).作为一种通用框架,MSCAS框架可对大多数模型设计类SFF方法与深度学习类SFF方法进行嵌入,进而实现方法在噪声鲁棒性、尺度敏感性方面的性能提升,本文主要贡献有2方面:

1)提出了框架MSCAS,该框架主要分为多尺度



(a) RR算法的噪声和尺度鲁棒性结果



(b) FV算法的噪声和尺度鲁棒性结果

Fig. 1 Impact of different scales and noise on the reconstruction results of typical SFF methods

图1 不同尺度与噪声对典型SFF方法重建结果的影响

分解与跨尺度代价聚合2个关键核心步骤. 多尺度分解有助于增加场景的深度表征信息选择空间；跨尺度代价聚合不仅加入了尺度间与尺度内的双重深度信息约束，而且在整个聚合过程中充分考虑图像序列中各相邻图像之间的聚焦趋势变化关系，有效提升场景重建精度的同时加速重建结果的鲁棒性收敛.

2）MSCAS作为一种通用框架，现有的大多数模型设计类SFF方法与深度学习类SFF方法均可以嵌入MSCAS框架，实现不同类型SFF重建方法性能的提升.

1 相关工作

1.1 多聚焦图像3维形貌重建原理

多聚焦图像3维形貌重建SFF的原理是利用图像中聚焦区域反推相对深度的一种方法. 该方法通过调节成像端与场景之间的相对距离，获得场景不同焦距的图像序列；然后采用聚焦测量（focus measure, *Fm*）函数对图像序列进行聚焦水平评价，形成聚焦度量堆栈；最后，通过聚合所有聚焦度量堆栈最大值所在索引，即可得场景的3维重建结果. 其过程如式（1）所示.

$$D_i = \arg\max_{1\leqslant n\leqslant N}\{Fm_{i,n}\}, \tag{1}$$

其中图像序列总数为N，单帧图像大小为$H\times W$，$Fm_{i,n}$表示第n张图像中第i个像素索引位置的焦点测量结果，D_i表示第i个像素索引位置对应的深度信息，最后对深度结果进行3维空间映射可得到待测场景的3维结构信息[12].

1.2 多聚焦图像3维形貌重建研究进展

多聚焦图像3维形貌重建可分为模型设计重建与深度学习重建两大类方法.

模型设计重建方法按照处理流程可以分为聚焦测量与后处理2个关键步骤[13]. 其中，聚焦测量主要用来评价图像序列中的聚焦水平，进而获得场景的初始深度图，该过程根据聚焦信息判断的差异可以分为空间域、频率域、统计类3类. 空间域聚焦测量评价方法主要通过聚焦测量算子与图像的卷积操作判断图像的聚焦信息，代表性的聚焦测量算子包括Laplace算子及其变体[14]、梯度类算子[15]、可变形算子等[16]；频率域聚焦测量方法主要将图像变换至频域空间，然后利用高频信息判断图像的聚焦水平，代表性方法包括非降采样小波变换[17]、非降采样剪切波变换[18]、Curvelet变换[19]等；统计类聚焦测量方法主要通过统计学指标计算图像的聚焦程度，代表性方法包括基于熵的聚焦测量算子和本征值等[20]. 由于聚焦测量过程形成的初始深度会包含较多的深度不连续信息，后处理主要针对初始深度图的修复展开，典型的方法包括罚函数法、正则化、迭代修复法[21]等. 由于模型设计类聚焦判别重建方法依赖于聚焦测量方法设计的鲁棒性与有效性，其对聚焦信息判断的误差可能会传播至后处理过程，导致重建结果无法收敛. 因此，深度学习类重建方法正逐渐引起研究者们的关注.

深度学习类重建方法首先需要构建多聚焦图像序列与深度真值数据集，然后通过训练端到端的深度网络模型对场景的深度信息进行预测. 因此，现有的深度学习类重建方法主要包括数据集构建与深度网络模型设计2类.

由于多聚焦图像3维形貌重建领域缺乏标准的真实数据集，现有的数据集大多通过其他任务的数据移植而来. 如LF_4D_PSF数据集[22]来自于4维光场重建领域，该数据集包括人为设计的多组室内场景，数据集包含场景的全聚焦图像与深度图像，通过深度图像的点扩散函数与全聚焦图像的卷积操作获得多幅不同聚焦水平的图像序列；Middlebury_FS数据集[23]来源于多目立体匹配领域，该数据集通过RGB相机与结构光扫描仪获取场景图与视差图；Flying-Things3D数据集[24]来源于光流估计领域，主要通过深度网络将光流估计扩展至视差估计；FoD500数据集[25]来自于虚焦深度估计领域，但该数据无法有效覆盖场景完整的深度范围. 基于上述高质量的合成数据集，借助相关网络模型可实现多聚焦图像的3维形貌结构的预测与推理.

在网络模型设计方面，DDFF模型主要提出一种

单纯的端到端的卷积神经网络，并利用构建的 DDFF-12 数据集进行训练与预测[26]；AiF 在网络模型中设计了一个共享的中间注意力模块，可实现场景中全聚焦图像与深度图像的同时推理[27]；FV 与 DFV 网络模型重新审视了 SFF 重建问题的本质，构建了聚焦差分模块实现场景的深度信息预测[11]；DFW 模型提出一种对齐网络实现了错位图像序列的配准，然后设计了聚焦信息检测模块获取场景的焦点信息[22]；DSFFNet 通过设计可变形卷积与局部-全局关系耦合模块提升模型的全局聚焦信息捕获能力[28]. 因此，深度学习类 SFF 重建方法致力于对网络模型进行有效设计，进而提升多聚焦图像序列中聚焦体积特征刻画的准确性.

1.3 代价聚合

代价聚合源于双目立体匹配问题[29]，通常情况下双目立体匹配算法首先需要进行代价计算，即在一定的视差搜索范围内计算待匹配像素与候选像素之间的相关性，这种相关性通过代价值体现，由于代价计算过程通常仅考虑局部信息，极易受到噪声与弱纹理区域的影响. 因此代价聚合则是基于一些特定假设(如相邻像素的视差值一致)使得聚合后的代价值能够更为准确地反映像素之间的相关性，其原理是将一个位置的加权响应传播到附近位置，该权重由参考图像像素之间的先验知识确定[30]. 而在 SFF 重建中，对不同聚焦水平图像序列的聚焦测量函数 *Fm* 可以视为一种代价量变体，代价聚合过程亦可对 *Fm* 值进行优化.

代价聚合可分为局部方法和全局方法. 局部代价聚合方法的一个例子是边缘保持过滤器. 由于深度图中的异质过渡区域通常对应于场景图像中的边缘，因此可以使用双边滤波器和图像引导滤波器[31] 作为代价聚合方法. 这些滤波器使用局部窗口优化代价量，同时保留引导图像中的边缘，由于局部代价聚合方法中仅仅利用了局部特征进行聚合，得到的视差对全局特征效果较差. 全局代价聚合方法将全局特征整合为一个能量函数，能量函数主要对整幅视差图进行能量评估，最小化能量函数可得到最终视差图，但其计算复杂度高且效率低下[32]. 上述方法[31-32] 基本仅考虑单一尺度情况，由于单一尺度的代价聚合容易受到噪声的影响，特别是对弱纹理或无纹理区域的聚合，可能会导致视差计算过程中误差的传播. 而多尺度代价聚合[33] 在捕获不同尺度的多层次特征时，能够对不同尺度的图像局部信息进行融合与筛选[34]，由于现有的多尺度代价聚合方法在分解过程存在降采样操作，且这种操作在特征融合过程中可能会出现信息损失，这无法直接应用于 SFF 重建过程，因为 SFF 重建严重依赖于图像的细节特征，任何降采样操作都可能损失一些关键深度表征信息. 因此，设计非降采样操作的多尺度代价聚合并结合 SFF 重建本身具有的序列关联特点，有望解决目前 SFF 重建方法面临的深度信息有效聚合收敛的问题.

2 多尺度代价聚合 SFF 重建框架

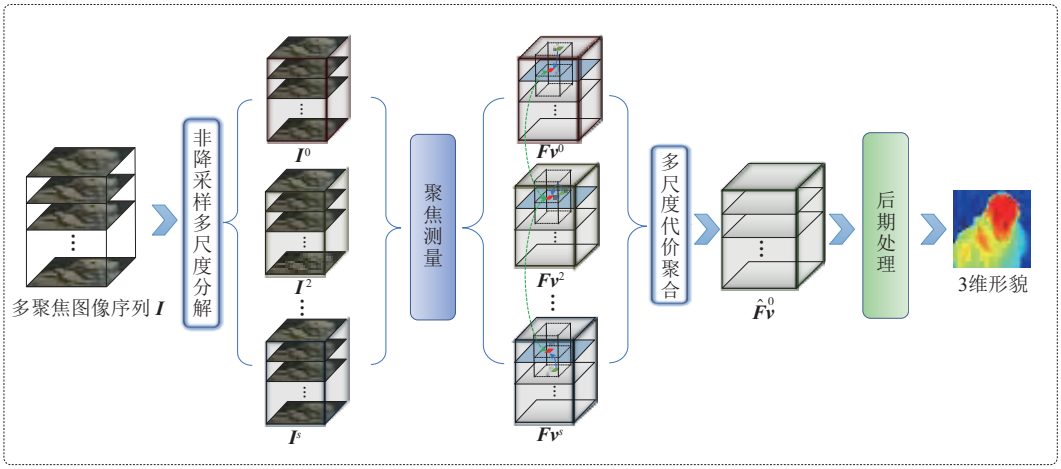
2.1 框架概述

研究表明无论是模型设计类 SFF 方法还是深度学习类 SFF 方法，均需通过聚焦体积(focus volume, ***Fv***)导出场景最终的深度结果[28]. 因此 ***Fv*** 的准确性是决定重建结果的关键. 现有的 ***Fv*** 的计算方法和设计模式大多从单一尺度进行评价，无法有效利用多尺度之间的有效信息. 尽管一些研究也采用了多尺度聚焦评价的方式[19]，但这类方法缺乏图像序列之间关联关系的约束，容易导致深度信息的不连续.

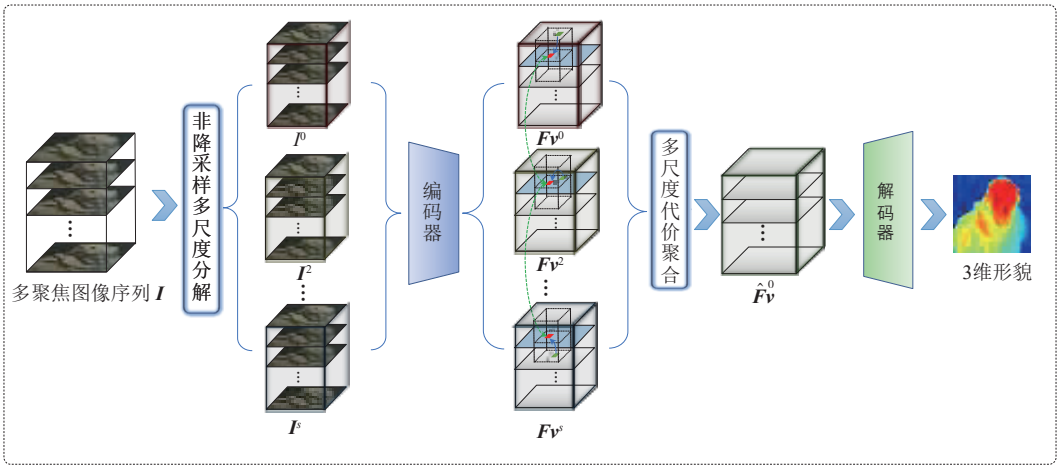
本文提出的 MSCAS 框架如图 2 所示. MSCAS 框架引入了非降采样的多尺度分解与多尺度代价聚合操作，这种扩展-聚合的模式不仅考虑了单一尺度下图像序列之间的关联关系，而且同时加入了不同尺度之间的内在联系，有助于深度表征信息的扩增与跨尺度深度信息的融合收敛. 可适配于现有主流的模型设计类 SFF 方法与深度学习类 SFF 方法中.

2.2 Atrous 多尺度分解

多尺度分解的主要目的是为了得到多聚焦图像序列的多尺度表达，由于典型的多尺度分解方法大都涉及图像序列的降采样操作，这会导致后续聚合过程中深度表征信息的缺失. Atrous 多尺度分解，因其滤波器上采样操作替代图像的下采样操作，可保证分解前后的信息在尺度上的一致性，广泛应用于图像融合[35] 和图像去噪[36] 等领域. 目前，已有研究利用 Atrous 多尺度分解算法应用于多聚焦图像序列融合[7]，但该研究采用 Atrous 多尺度分解仅为获得质量较好的融合图像来引导深度信息，未考虑不同尺度对深度信息本身存在的影响. 除此之外，也有研究利用 Atrous 多尺度分解算法解决 SFF 问题[21]，但该研究仅采用了其中单一尺度的信息进行深度回溯，没有充分利用多尺度信息蕴含的丰富的深度线索. 因此，本文首先对多聚焦图像序列进行 Atrous 多尺度分解，可保证分解后的图像序列与原始图像序列的大小一致，然后在此基础上充分挖掘不同尺度中的深度线索.



(a) 模型设计类SFF方法的多尺度代价聚合示意图



(b) 深度学习类SFF方法的多尺度代价聚合示意图

Fig. 2 Schematic diagram of multi-scale cost aggregation in model-based SFF and deep learning-based SFF methods reconstruction framework

图2 多尺度代价聚合在模型类SFF方法与深度学习类SFF方法的重建框架示意图

定义 1. 给定输入图像序列中图像的大小为$H\times W$，则N幅图像序列组成的多聚焦图像序列可表示为

$$\boldsymbol{I}=(I_i(x,y))_{i=1}^{N}, \tag{2}$$

其中$1\leqslant x, y\leqslant N$.

定义 2. 给定一个S级塔式分解滤波器组$\boldsymbol{H}_s(z)$，$0\leqslant s\leqslant S$，则多聚焦图像序列$\boldsymbol{I}$在$S+1$个尺度的低频子图$\boldsymbol{I}^s$可表示为

$$\boldsymbol{I}^s=\boldsymbol{H}_s(z)*\boldsymbol{I},\ 0\leqslant s\leqslant S, 1\leqslant z\leqslant N, \tag{3}$$

其中$\boldsymbol{H}_s(z)$为滤波器组$\{\boldsymbol{H}_s\}$的z变换，表达式为

$$\boldsymbol{H}_s(z)=\prod_{j=0}^{s}\boldsymbol{H}_0(z^{2^j}),\ 0\leqslant s\leqslant S, 1\leqslant z\leqslant N, \tag{4}$$

其中$\boldsymbol{H}_0(z)$为低通分解滤波器.

定义2中后一级滤波器$\boldsymbol{H}_0(z^{2^{j+1}})$由前一级滤波器$\boldsymbol{H}_0(z^{2^j})$按照采样矩阵$2\boldsymbol{E}$($\boldsymbol{E}$为2阶单位矩阵)进行上采样得到，因此经过塔式滤波器后的多尺度图像序列$\boldsymbol{I}^s$与原图像序列$\boldsymbol{I}$具有相同大小，这种不同尺度对齐的方式不仅可减少后续多尺度代价聚合过程中复杂的权重划分问题，而且可有效保留不同尺度之间同一位置的深度表征信息，进而有助于增强深度信息判断的准确性.

2.3 多尺度代价聚合

多尺度代价聚合是MSCAS框架的核心，该模块的主要目的是在不同尺度上对聚焦体积进行逐尺度逐层聚合，单一尺度聚合过程在原有单帧图像的邻域范围的基础上加入了图像序列之间的关联关系.跨尺度聚合过程则考虑了不同尺度的权重划分问题.

定义 3. 给定一个聚焦测量(focus measure, Fm)方法，该方法可建立图像序列$\boldsymbol{I}$与聚焦体积$\boldsymbol{Fv}$之间的映射关系：

$$\boldsymbol{Fv}=Fm(\boldsymbol{I}), \tag{5}$$

其中Fm可以是模型设计类方法的聚焦测量算子或者深度学习类方法的编码过程，对于单个像素i位于

(x_i,y_i)(x_i,y_i表示像素位置)而言，$Fv(i,l)$ 表示第l幅图像中(x_i,y_i)位置的聚焦测量结果.

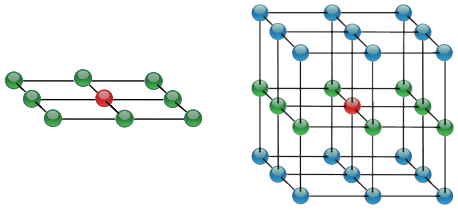
由于 SFF 方法对多聚焦图像序列采用单帧图像独立聚焦评价的模式，而在实际场景中广泛存在稀疏纹理和低对比度区域的情形，这可能导致聚焦体积 $\boldsymbol{Fv}$ 中包含较多的噪声信息. 受加权最小二乘法去噪的启发，引入代价聚合可对 $\boldsymbol{Fv}$ 中的噪声进行有效去除，去噪后的聚焦体积$\widetilde{\boldsymbol{Fv}}$表示为

$$\widetilde{Fv}(i,l)=\arg\min_{z}\frac{1}{Z_i}\sum_{j\in N_i}W(i,j)\|z-Fv(j,l)\|^2, \tag{6}$$

其中N_i表示像素i的邻域范围，$W(i,j)$表示像素i与像素j之间的相似性度量，Z_i为归一化常量，其表达式为

$$Z_i=\sum_{j\in N_i}W(i,j). \tag{7}$$

由于代价聚合过程本身来自于双目立体匹配，因此典型的代价聚合过程仅考虑单幅图像邻域之间的联系. 由于多聚焦图像序列之间存在较为密切的关联关系，因此本文提出的 MSCAS 框架中将相似性度量过程扩展至序列之间，有助于解决低对比度或者稀疏纹理区域的有效聚合问题. 具体邻域设计的差异如图 3 所示.



(a) 现有方法邻域 (b) 本文框架邻域

Fig. 3 Differences in neighborhood design between existing cost aggregation methods and MACAS framework

图 3 现有代价聚合方法与 MACAS 框架在邻域设计的差异

式(6)的解可通过式(8)得到：

$$\widetilde{Fv}(i,l)=\frac{1}{Z_i}\sum_{j\in N_i}W(i,j)Fv(j,l). \tag{8}$$

研究表明 2 维图像中的颜色纹理信息通常与深度图中同质和异质区域呈现一定相关性. 受引导滤波过程启发，像素间的相似性度量$W(i,j)$采用式(9)计算.

$$W(i,j)=\frac{1}{|\omega|^2}\sum_{k:(i,j)\in\omega_k}\left(1+\frac{(I_i-\mu_k)(I_j-\mu_k)}{\sigma_k^2+\varepsilon}\right), \tag{9}$$

其中μ_k和σ_k是图像序列$\boldsymbol{I}$以第k个像素为中心，以$r\times r\times r$立方体范围为窗口的均值和方差. I_i和I_j表示图像序列中对应位置的像素值，$|\omega|$表示该窗口中的像素数，ε是平滑度参数，在本文中参考文献 [30] 设置$\varepsilon=0.01^2$.

当加入尺度信息后，对于任一尺度$s\in\{0,1,2,\cdots,S\}$，其可通过式(10)独立求解.

$$\widetilde{Fv}^s(i^s,l^s)=\arg\min_{\{z^s\}_{s=0}^S}\frac{1}{Z_{i^s}^s}\sum_{j\in N_{i^s}}W(i^s,j^s)\|z^s-Fv^s(j^s,l^s)\|^2. \tag{10}$$

根据式(8)可以计算得到式(10)的解为：

$$\forall s,\widetilde{Fv}^s(i^s,l^s)=\frac{1}{Z_{i^s}^s}\sum_{j\in N_{i^s}}W(i^s,j^s)C(j^s,l^s). \tag{11}$$

由于不同尺度之间聚焦体积存在信息互补的情形，因此为有效利用不同尺度之间的联系，通过引入正则化项可实现不同尺度之间的约束. 式(10)则变为如下优化问题：

$$\widehat{Fv}^s(i^s,l^s)=\arg\min_{\{z^s\}_{s=0}^S}\left(\frac{1}{Z_{i^s}^s}\sum_{j\in N_{i^s}}W(i^s,j^s)\|z^s-Fv^s(j^s,l^s)\|^2+\lambda\sum_{s=1}^{S}\left\|z^s-z^{s-1}\right\|^2\right). \tag{12}$$

通过对式(12)求导可得

$$\frac{\partial\widehat{\boldsymbol{Fv}}^s}{\partial z^s}=\frac{2}{Z_{i^s}^s}\sum_{j^s\in N_{i^s}}W(i^s,j^s)(z^s-Fv^s(j^s,l^s))+2\lambda\left(z^s-z^{s-1}\right)-2\lambda\left(z^{s+1}-z^s\right)=2\left(-\lambda z^{s-1}+(1+2\lambda)z^s-\lambda z^{s+1}-\widetilde{Fv}^s(i^s,l^s)\right). \tag{13}$$

令$\dfrac{\partial\widehat{\boldsymbol{Fv}}^s}{\partial z^s}=0$，即有

$$\widetilde{Fv}^s(i^s,l^s)=-\lambda z^{s-1}+(1+2\lambda)z^s-\lambda z^{s+1}. \tag{14}$$

同理上述结果适用于尺度范围 0~S 的情形，因此可以建立 $\widehat{\boldsymbol{Fv}}^s$与$\widetilde{\boldsymbol{Fv}}^s$的联系，即

$$\boldsymbol{A}\widehat{\boldsymbol{Fv}}^s=\widetilde{\boldsymbol{Fv}}^s, \tag{15}$$

其中矩阵$\boldsymbol{A}$是由不同尺度构成的三角矩阵，当最大尺度$S=2$时，其表达式为

$$\boldsymbol{A}=\begin{pmatrix}1+\lambda & -\lambda & 0\\ -\lambda & 1+2\lambda & -\lambda\\ 0 & -\lambda & 1+\lambda\end{pmatrix}, \tag{16}$$

且其必然存在逆矩阵$\boldsymbol{A}^{-1}$，则可得到

$$\widehat{\boldsymbol{Fv}}^s=\boldsymbol{A}^{-1}\widetilde{\boldsymbol{Fv}}^s. \tag{17}$$

最后将所有尺度汇总到最精细尺度$\widehat{\boldsymbol{Fv}}^0$：

$$\widehat{Fv}^0\left(i^0,l^0\right)=\sum_{s=0}^{S}\boldsymbol{A}^{-1}(0,s)\widetilde{Fv}^s(i^s,l^s). \tag{18}$$

通过式(19)得到多尺度代价聚合后聚焦体积的深度结果：

$$\boldsymbol{D}_{i^0}=\arg\max_{l^0}\left\{P\left(\widehat{Fv}^0\left(i^0,l^0\right)\right)\right\}, \tag{19}$$

其中$\boldsymbol{D}$为 3 维形貌重建结果，由于i^0为最精细尺度，所以有$i^0=i$，$l^0=l$. $P(\cdot)$为模型设计类方法的后处理

过程或者为深度学习类方法解码过程.

通过上述多尺度代价聚合过程，MSCAS框架不仅在特定尺度考虑了图像序列间的关联关系，还将跨尺度之间的信息约束考虑在内，进而从水平的单一尺度与垂直的跨多尺度2个层面提升深度信息判断的准确性与抗噪鲁棒性.

2.4 MSCAS算法流程

算法1. 多尺度代价聚合SFF重建框架MSCAS.

输入：多聚焦图像序列$\boldsymbol{I}$；

输出：根据式(19)得到场景的3维形貌重建结果$\boldsymbol{D}$.

① 图像序列$\boldsymbol{I}$进行Atrous多尺度分解，得到$\boldsymbol{I}^s$，$0\leqslant s\leqslant S$；

② 对任意尺度使用聚焦评价得到聚焦体积$\boldsymbol{Fv}^s$，$0\leqslant s\leqslant S$；

③ 根据式(11)对所有尺度的$\boldsymbol{Fv}^s$进行代价聚合得到$\widetilde{\boldsymbol{Fv}}^s$；

④ 根据式(18)计算多尺度代价聚合后的聚焦体积$\widehat{\boldsymbol{Fv}}^0$.

3 实验及结果分析

3.1 实验设置

本文实验可分为2部分：模型设计类SFF重建方法与深度学习类SFF重建方法. 其中模型设计类SFF重建方法均使用Matlab中的图像处理与计算机视觉工具包实现，深度学习类SFF方法则使用PyTorch框架实现，运算环境为Ubuntu 18.04.6 LTS, 62.5 GB DDR4 DDIM, Intel® Xeon® W-2255 CPU @ 3.70 GHz×20, NVIDIA GeForce RTX 4090, 24 GB显存. 模型设计类SFF重建各方法均采用原算法默认参数. 深度学习类SFF重建方法中所有模型的训练设置相同，使用Adam优化器，初始学习率设置为10^{-4}，训练100轮次，批处理大小设置为4.

3.2 数据集

本文采用SFF重建方法常用的4种多聚焦图像3维形貌重建基准数据集LF_4D_PSF[22], Middlebury_FS[23], FlyingThings3D[24], FoD500[25]. 各数据集统计信息如表1所示.

Table 1 Statistic Information of Datasets

表1 数据集统计信息

数据集	图像来源	场景数	图像大小
LF_4D_PSF	合成场景	24	512×512×30
Middlebury_FS	真实场景	33	2 964×1 988×15
FlyingThings3D	合成场景	1 000	960×540×15
FoD500	合成场景	500	256×256×5

LF_4D_PSF数据集来源于4维光场重建领域，包含24个模拟场景，每个场景提供了场景图与深度真值. 每个场景的深度真值通过点扩散函数映射至场景图以实现30幅不同聚焦水平的图像序列.

Middlebury_FS数据集来源于立体匹配3维重建领域，包含33个静态室内场景，提供了相机标定参数和结构光扫描仪获取的场景真值视差图. 每个场景包含15幅不同聚焦水平的堆栈，场景包括重复结构、遮挡、细线状物体以及无纹理区域.

FlyingThings3D数据集来源于光流估计领域，包含1 000幅随机生成的飞行物体及3D轨迹，该数据集将卷积网络光流估计的概念扩展到视差估计，每个场景包含15幅不同聚焦水平的堆栈.

FoD500数据集来源于单目深度估计领域，包含500个场景，由于该数据集采用散焦模式生成，因此每个场景包含的5幅不同聚焦水平的堆栈无法有效覆盖场景全部景深范围，因此本文实验仅比较其量化指标.

3.3 对比方法

本文分别选择4种模型设计类SFF方法与4种深度学习类SFF方法进行实验分析. 各对比方法的参数设置见表2.

Table 2 Information Summary of Comparative SFF Methods

表2 对比SFF方法信息汇总

对比方法	归属类型	期刊或会议/年份	参数设置
SML	模型设计类SFF	PR/2013	T_1=7，*step*=1，*N*=2
GD	模型设计类SFF	PR/2013	*Sigma*=2，*hsize*=5
RDF	模型设计类SFF	TIP/2020	*Tmad*=0.1，*Tbokeh*=0.15，*Fm* (r_1=1，r_2=2，r_3=5)
RR	模型设计类SFF	TIP/2021	*lambda*=0.3，*alpha*=0.1，*beta*=1.5，*gamma*=2.5，*itr*=8，*nei*=2
DFV	深度学习类SFF	CVPR/2022	Adam（$\beta_1=0.9$，$\beta_2=0.999$），*alpha* ∈{8/15，4/15，2/15，1/15}
FV	深度学习类SFF	CVPR/2022	Adam（$\beta_1=0.9$，$\beta_2=0.999$），*alpha* ∈{8/15，4/15，2/15，1/15}
DFW	深度学习类SFF	ECCV/2022	Adam（$\beta_1=0.9$，$\beta_2=0.99$）
AiF	深度学习类SFF	ICCV/2021	Adam（$\beta_1=0.9$，$\beta_2=0.999$）

1)4种模型设计类SFF方法

①修正拉普拉斯和函数(sum-modified Laplacian, SML)[13]. SML作为聚焦度量算子，主要通过提取图像边缘判定图像的清晰区域，该方法对亮度变化不敏感，可得到明显聚焦峰值，表达式为

$$Fm_{\mathrm{SML}}(x,y)=\sum_{u=-k}^{k}\sum_{v=-k}^{k}\left|I_{xx}(x+u,y+v)\right|+\left|I_{yy}(x+u,y+v)\right|, \tag{20}$$

其中I_{xx}和I_{yy}代表图像在x和y方向上的2阶偏导数，k表示模板的半径. u和v表示模板中的像素位置相对于当前像素(x,y)的偏移量.

②高斯导数(Gaussian derivative, GD)算子[13]. GD算子通过高斯平滑有效地抑制噪声，由于拥有不同大小的高斯核，具备较好的尺度不变性，表达式为

$$Fm_{\mathrm{GD}}(x,y)=-\frac{1}{2\pi\sigma^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{21}$$

其中σ为高斯函数的宽度标准差.

③环差滤波器(ring difference filter, RDF)[16]. 该算子通过设计不同半径的环形模板，将多个环形模板得到的局部聚焦度量值进行线性加权平均以获取图像的局部和全局信息，进而得到综合聚焦度量值. 其表达式为

$$Fm_{\mathrm{RDF}}=\begin{cases}\dfrac{2}{\pi R_1^2}, & |x-y|<R_1,\\ -\dfrac{2}{\pi\left(R_3^2-R_2^2\right)}, & R_2\leqslant|x-y|\leqslant R_3,\end{cases} \tag{22}$$

其中R_1,R_2,R_3分别为环半径.

④鲁棒正则化(robust regularization, RR)[10]. RR方法采用体积约束进行正则化，将3维形貌重建问题转化为一个罚函数优化问题，并通过最小化能量函数来实现深度估计，其表达式为：

$$Fm_{\mathrm{RR}}=\sum_{p\in\Omega}(u_p-Fm_{\mathrm{ML}})^2+\lambda\Theta(u), \tag{23}$$

其中u_p为数据保真项，Fm_{ML}为修正的拉普拉斯聚焦测量算子，$\Theta(u)$为正则化项，λ控制u_p和$\Theta(u)$两项的相对重要性.

2)4种深度学习类SFF方法

①深度差分聚焦体积(deep differential focus volume, DFV)网络[11]. DFV旨在通过使用不同焦距上的堆叠特征计算一阶导数，进而捕获焦点和上下文信息辅助进行焦点分析.

②聚焦体积(focus volume, FV)网络[11]. FV将4维聚焦体积引入SFF任务，通过最大值自变量点集计算位置概率信息，具有高效实时等特点，可以快速响应图像中的轮廓和边缘信息.

③DFW(depth from wild)[23]. DFW网络提出了可学习的基于上下文图像对齐方法，能够对聚焦图像进行有效对齐. 引入了锐化区域检测模块减少弱纹理区域中细微焦点变化引起的模糊歧义，同时设计了一种高效下采样模块提取焦点信息流.

④AiF(all-in-focus)[27]. AiF方法连接了有监督和无监督学习，AiF提出了一种共享网络，可从不同聚焦图像序列中利用一个中间的注意力机制同时估计深度和全焦图像.

3.4 评价指标

本文在重建结果评价方面采用4种评价指标：均方根误差(root mean squared error, *RMSE*)、结构相似度(structural similarity index measure, *SSIM*)、相关系数(correlation coefficient, *Corr*)、峰值信噪比(peak signal-to-noise ratio, *PSNR*)，其公式为：

$$\begin{cases}RMSE=\sqrt{\dfrac{1}{M}\sum\limits_{i\in M}(GT_i-D_i)^2},\\ SSIM=\dfrac{(2\mu_{GT}\mu_D+C_1)(2\sigma_{GT,D}+C_2)}{(\mu_{GT}^2+\mu_D^2+C_1)(\sigma_{GT}^2+\sigma_D^2+C_2)},\\ Corr=\dfrac{\sum\limits_{i\in M}(D_i-\mu_D)(GT_i-\mu_{GT})}{\sqrt{\sum\limits_{i\in M}(D_i-\mu_D)^2}\sqrt{\sum\limits_{i\in M}(GT_i-\mu_{GT})^2}},\\ PSNR=10\times\lg\left(\dfrac{max\times M}{\sum\limits_{i\in M}(GT_i-D_i)^2}\right),\end{cases} \tag{24}$$

其中GT代表场景真值深度图，D代表预测深度图，M代表深度图的总像素数，μ代表深度均值，σ代表深度标准差，max代表图像中像素比较的最大范围.

3.5 正则化参数λ的选择

由式(12)可知，正则化参数λ可调整不同尺度之间权重的变化，λ越小表明最精细尺度 ***Fv*** 起主要作用，随着λ的增加，其他尺度的权重逐渐增加. 为客观评价本文框架MSCAS对不同方法性能的提升，本节实验将正则化参数λ的范围设置为0~2，观察不同方法在LF_4D_PSF数据集中平均*RMSE*指标的变化情况.

图4表明所有方法随着λ逐渐增大平均*RMSE*值

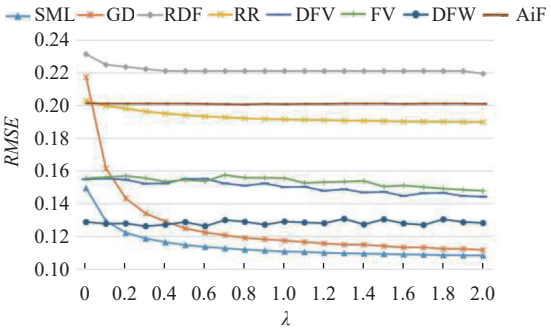


Fig. 4 Experimental setting for different regularization parameters

图4 不同正则化参数设置实验

出现降低，在$\lambda \geqslant 1$时 *RMSE* 下降幅度逐渐放缓. 因此，针对不同方法设计不同的λ值会使得方法的性能得到有效提升，为兼顾不同方法间的性能，同时便于分析比较 MSCAS 框架对各方法的提升效果，本文推荐$\lambda=1$作为各方法的正则化参数值.

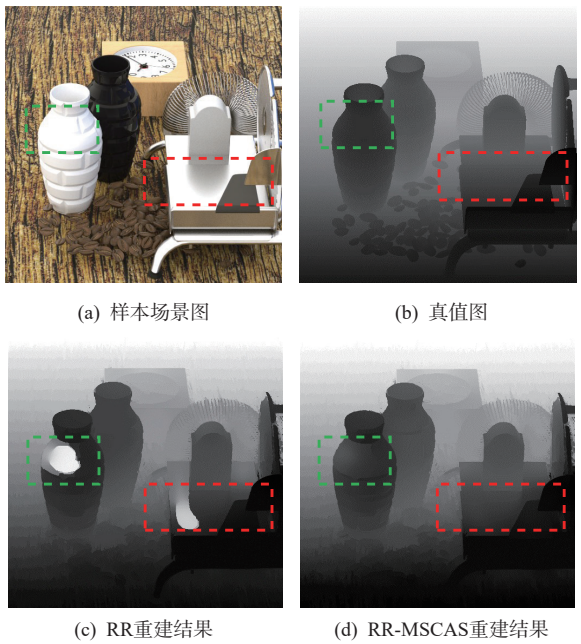
3.6 稀疏区域样本的定性分析

本节选择具有稀疏区域的样本验证 MSCAS 框架的有效性. 由图 5(a)(b)可知，场景中小矩形框中的白色瓶体因为光照原因导致了表面具有稀疏纹理的特性，而大矩形区域由于场景物体表面的金属高反光特性也产生了弱纹理区域. 由图 5(c)(d)可以看出，RR 方法因为未考虑图像序列间聚焦信息的变化趋势，导致其无法对这 2 部分区域进行有效重建，而嵌入本文框架后的 RR-MSCAS 方法则充分考虑了尺度内序列间关联关系与跨尺度间深度表征信息的互补特性，可对这 2 类区域实现有效重建.

3.7 尺度鲁棒性对比实验分析

为验证本文 MSCAS 框架在不同尺度下的鲁棒性，本节选择对场景的结构信息保留程度较好的平均 *SSIM* 值作为评价指标，并以 LF_4D_PSF 数据集为例，将该数据集扩展为 512×512，256×256，128×128，64×64 这 4 种尺度，分别在 8 种 SFF 重建方法进行实验.

图 6 中不带网点的 S*x*OA($x\in\{1,2,3,4\}$分别表示上述 4 种不同尺度)代表未嵌入 MSCAS 的原方法，带网点的 S*x*MSCAS 代表嵌入 MSCAS 后的 SFF 重建方法. 结果可以发现不同尺度中嵌入 MSCAS 框架的 S*x*MSCAS 方法的平均 *SSIM* 值有显著提升. 具体而言，嵌入 MSCAS 框架后方法在 512×512 尺度上平均 *SSIM* 提升了 15.04 个百分点，在 256×256 尺度上平均提升了 27.88 个百分点，在 128×128 尺度上平均提升了 47.14 个百分点，在 64×64 尺度上平均提升了 64.87 个百分点. 实验结果呈现出尺度越低，性能提升越明显的趋势. 因此，引入 MSCAS 框架可有效提升不同类型 SFF 重建方法的尺度鲁棒性.



(a) 样本场景图　(b) 真值图　(c) RR重建结果　(d) RR-MSCAS重建结果

Fig. 5 Comparison between RR method and RR-MSCAS method in sparse texture samples

图 5 RR 方法与 RR-MSCAS 方法在稀疏纹理样本的比较

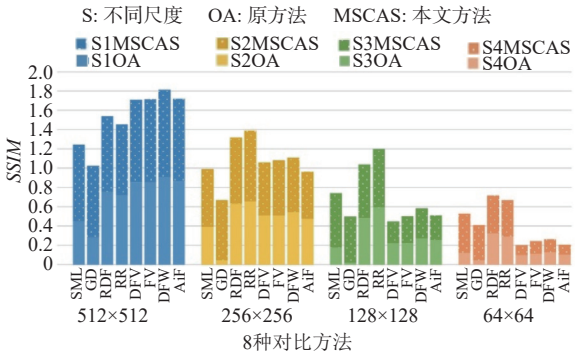


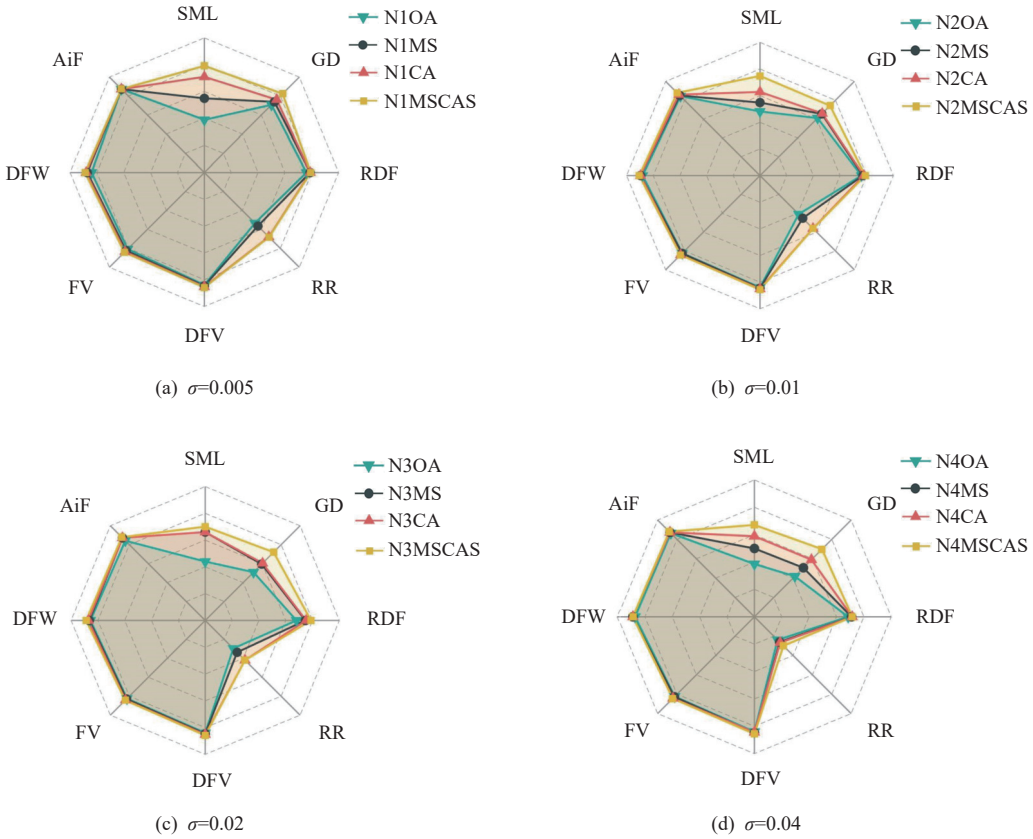
Fig. 6 Comparison between the original method and the embedded MSCAS framework at different scales

图 6 原方法与嵌入 MSCAS 框架后在不同尺度下的对比

3.8 噪声鲁棒性对比实验分析

为验证本文 MSCAS 框架的噪声鲁棒性，本节以 LF_4D_PSF 数据集为例，分别对该数据集加入$\sigma=0.005$，$\sigma=0.01$，$\sigma=0.02$，$\sigma=0.04$这 4 种不同强度的高斯白噪声. 然后以平均 *SSIM* 测试 8 种不同算法的噪声鲁棒性. 具体从以下 4 个角度进行实验分析，分别是加入噪声的原方法(N*x*OA，$x\in\{1,2,3,4\}$分别表示上述 4 种不同强度噪声)、仅加入多尺度分解的方法(N*x*MS)、仅加入跨尺度代价聚合方法(N*x*CA)和嵌入本文 MSCAS 框架的方法(N*x*MSCAS).

由图 7 可知，嵌入 MSCAS 框架后的各 SFF 方法 N*x*MSCAS 在不同强度噪声干扰下性能均优于 N*x*MS 方法、N*x*CA 方法和原方法 N*x*OA. 其中 N*x*CA 方法的贡献显著优于 N*x*MS 方法. 具体结果为：当噪声强度$\sigma=0.005$时(图 7(a))，N*x*MSCAS 方法、N*x*CA 方法和 N*x*MS 方法相比于 N*x*OA 方法的平均 *SSIM* 分别提升了 29.21 个百分点、23.20 个百分点、12.58 个百分点，在$\sigma=0.01$时(图 7(b))分别提升了 32.29 个百分点、18.63 个百分点、7.64 个百分点，在$\sigma=0.02$时(图



(a) σ=0.005 (b) σ=0.01

(c) σ=0.02 (d) σ=0.04

Fig. 7 Average *SSIM* of MSCAS framework components and original method under different noise levels

图 7 不同强度噪声下 MSCAS 框架各部分与原方法的平均 *SSIM*

7(c))分别提升了 30.87 个百分点、20.88 个百分点、11.69 个百分点，在 $\sigma=0.04$时(图 7(d))分别提升了 30.91 个百分点、20.72 个百分点、12.53 个百分点. 上述实验表明了嵌入 MSCAS 框架后各 SFF 重建方法具有了一定的噪声鲁棒性.

3.9 对比实验结果分析

本节主要从 3.4 节的 4 种客观评价指标和 3 维形貌重建的实际效果 2 个层面对本文提出的 MSCAS 框架进行定量与定性分析.

如表 3 所示，分别为将本文 MSCAS 框架嵌入 8 种不同类型的 SFF 重建方法中，并在 LF_4D_PSF, Middlebury_FS, FlyingThings3D, FoD500 这 4 种数据集中通过 *RMSE*, *SSIM*, *PSNR*, *Corr* 这 4 个评价指标的平均值进行统一量化分析. 由于深度学习类 SFF 重建方法将 *RMSE* 值做了归一化处理，为了便于统一分析比较，本节实验中将模型设计类 SFF 重建方法的 *RMSE* 指标也进行了归一化处理.

表 3 表明本文 MSCAS 框架对模型设计类 SFF 重建方法性能的提升显著优于深度学习类 SFF 重建方法. 这是由于深度学习类 SFF 重建方法基于上述数据集进行训练，可能会出现过拟合的情况，加之训练数据集场景样本多样性较少，如 FoD500 数据集中

Table 3 Performance Comparison of Eight SFF Methods Before and After Embedding MSCAS Framework Across Four Datasets

表 3 8 种 SFF 方法在嵌入 MSCAS 框架前后在 4 组数据集中的性能比较

数据集	方法类型	SFF 方法	*RMSE*		*SSIM*		*PSNR*		*Corr*	
			原方法	MSCAS	原方法	MSCAS	原方法	MSCAS	原方法	MSCAS
LF_4D_PSF	模型设计类方法	SML	0.149 6	**0.109 0**	0.446 9	**0.796 6**	18.366 8	**23.919 8**	0.869 6	**0.948 5**
		GD	0.222 3	**0.116 5**	0.283 2	**0.744 2**	14.545 2	**20.430 0**	0.633 6	**0.898 9**
		RDF	0.231 6	**0.219 9**	0.756 0	**0.784 2**	23.617 6	**24.291 1**	0.927 7	**0.940 4**
		RR	0.203 2	**0.200 8**	0.720 2	**0.732 6**	20.456 4	**21.668 9**	0.952 8	**0.953 3**
		平均值	0.201 7	0.161 6	0.551 6	0.764 4	19.246 5	22.577 5	0.845 9	0.935 3
		MSCAS-原方法	↓ 19.90%		↑ 38.58%		↑ 17.31%		↑ 10.56%	

表 3（续）

数据集	方法类型	SFF 方法	*RMSE*		*SSIM*		*PSNR*		*Corr*	
			原方法	MSCAS	原方法	MSCAS	原方法	MSCAS	原方法	MSCAS
LF_4D_PSF	深度学习类方法	DFV	0.153 7	**0.153 5**	0.855 0	**0.856 8**	31.330 6	**31.332 3**	0.972 4	0.972 6
		FV	0.154 9	**0.153 9**	0.857 0	**0.857 0**	31.274 6	**31.274 6**	0.972 2	**0.972 7**
		DFW	0.128 8	**0.126 7**	0.905 9	**0.907 5**	22.838 8	**22.940 7**	0.980 2	**0.980 7**
		AiF	0.201 4	**0.200 8**	0.860 9	**0.861 0**	24.789 7	**24.806 0**	0.961 0	**0.961 4**
		平均值	0.159 7	0.158 7	0.869 7	0.870 6	27.558	27.588 4	0.971 4	0.975 3
		MSCAS-原方法	↓0.61%		↑ 0.10%		↑ 0.11%		↑ 0.40%	
Middlebury_FS	模型设计类方法	SML	0.242 7	**0.158 7**	0.582 2	**0.835 7**	20.212 0	**28.516 0**	0.653 7	**0.784 2**
		GD	0.332 3	**0.231 5**	0.283 2	**0.744 2**	14.545 2	**20.430 0**	0.633 6	**0.898 9**
		RDF	0.217 3	**0.200 1**	0.756 0	**0.784 2**	23.617 6	**24.291 1**	0.927 7	**0.940 4**
		RR	0.237 3	**0.217 7**	0.699 3	**0.744 3**	20.419 2	**20.877 4**	0.730 3	**0.761 7**
		平均值	0.257 4	0.202 0	0.580 2	0.777 1	19.698 5	23.528 6	0.736 3	0.846 3
		MSCAS-原方法	↓ 21.52%		↑ 33.94%		↑ 19.44%		↑14.94%	
	深度学习类方法	DFV	0.207 2	**0.178 1**	0.625 9	**0.783 2**	16.533 2	**20.356 1**	0.627 7	**0.724 2**
		FV	0.231 5	**0.225 7**	0.719 5	**0.742 8**	18.548 0	**19.112 5**	0.647 8	0.645 3
		DFW	0.161 1	**0.160 4**	0.900 3	**0.905 6**	24.323 9	**24.718 0**	0.947 7	**0.957 3**
		AiF	0.186 3	**0.183 9**	0.831 6	**0.834 4**	15.635 1	**15.759 6**	0.647 0	**0.652 8**
		平均值	0.196 5	0.187 0	0.769 3	0.816 5	18.760 0	19.986 6	0.717 5	0.744 9
		MSCAS-原方法	↓ 4.83%		↑ 6.13%		↑ 6.54%		↑ 3.81%	
FlyingThings3D	模型设计类方法	SML	0.493 2	**0.423 3**	0.571 2	**0.767 6**	19.556 2	**22.239 6**	0.608 1	**0.777 1**
		GD	0.567 4	**0.494 1**	0.342 1	**0.685 5**	18.010 8	**21.030 6**	0.486 0	**0.690 8**
		RDF	0.429 6	0.468 8	0.808 4	**0.808 8**	22.551 9	**22.599 3**	0.808 2	0.799 5
		RR	0.359 5	**0.333 2**	0.677 7	**0.738 0**	21.314 8	**22.692 5**	0.661 7	**0.733 6**
		平均值	0.462 4	0.429 9	0.599 9	0.750 0	20.358 4	22.140 5	0.641 0	0.750 3
		MSCAS-原方法	↓ 7.04%		↑ 25.03%		↑ 8.75%		↑ 17.04%	
	深度学习类方法	DFV	0.167 1	**0.165**	0.721 1	**0.724 3**	17.413 6	**17.566 3**	0.763 7	0.765 9
		FV	0.178 5	**0.177 5**	0.658 9	**0.659 4**	11.103 7	**11.259**	0.778 3	0.779 6
		DFW	0.098 9	**0.098 2**	0.898 1	**0.900 5**	21.296 5	**21.383 9**	0.975 2	0.977 5
		AiF	0.122 1	**0.122 0**	0.930 3	0.930 7	20.044 1	**20.044 1**	0.815 7	**0.815 8**
		平均值	0.141 6	0.140 6	0.802 1	0.803 7	17.464 4	17.563 3	0.833 2	0.834 6
		MSCAS-原方法	↓ 0.68%		↑ 0.20%		↑ 0.56%		↑ 0.17%	
FoD500	模型设计类方法	SML	0.756 8	**0.655 6**	0.058 8	**0.290 7**	7.780 4	**7.425 5**	0.132 3	**0.402 1**
		GD	0.772 5	**0.767 2**	0.038 2	**0.456 9**	7.442 0	**9.026 2**	0.086 3	**0.224 5**
		RDF	0.651 7	**0.649 2**	0.411 9	**0.451 4**	8.039 4	**8.752 1**	0.467 5	**0.499 0**
		RR	0.758 0	**0.539 1**	0.210 3	**0.449 3**	13.915 4	**15.225 0**	0.203 6	**0.311 4**
		平均值	0.734 8	0.652 8	0.179 8	0.412 1	9.294 3	10.107 2	0.222 4	0.359 3
		MSCAS-原方法	↓ 11.16%		↑ 129.19%		↑ 8.75%		↑ 61.52%	
	深度学习类方法	DFV	0.166 7	**0.166 7**	0.906 4	0.906 4	31.047 2	**31.048 2**	0.854 8	**0.854 9**
		FV	0.166 7	**0.166 7**	0.906 4	**0.906 4**	31.047 5	**31.047 9**	0.854 8	**0.854 9**
		DFW	0.085 9	**0.085 7**	0.925 1	**0.925 2**	24.253 9	**24.254 1**	0.920 7	**0.921 3**
		AiF	0.150 9	**0.150 7**	0.910 4	**0.910 7**	30.203 6	**30.203 8**	0.874 2	**0.874 3**
		平均值	0.142 5	0.142 4	0.912 0	0.912 1	29.138 0	29.138 5	0.876 1	0.876 4
		MSCAS-原方法	↓ 0.07%		↑ 0.01%		↑ 0.00%		↑ 0.02%	

注：黑体数值表示性能提升的指标.

堆栈数仅为 5, 除此之外, *RMSE* 指标均做归一化处理, 所以体现在性能指标上的变化也最小.

图 8 选择了 LF_4D_PSF 数据集中的 2 个样本进行主观分析. 从矩形区域可以看出, 本文 MSCAS 框架对模型设计类 SFF 重建方法的提升效果明显, 特别是较好地聚合了 GD 方法原始稀疏的重建结果; 除此之外, MSCAS 框架可对 RR 方法的错误深度点进行有效聚合; 最后本文 MSCAS 框架对深度学习类 SFF 算法在边缘细节方面有一定提升.

图 9 为 Middlebury_FS 数据集中选择的场景, 且该数据集中的深度真值本身存在一些深度缺失区域. 从矩形区域可以看出, MSCAS 框架不仅对深度信息缺失区域有一定的修复效果, 而且对 RDF 和 RR 方法的细节保留和噪声去除方面有一定效果, 特别是对深度学习类的 FV 方法的有显著提升, 如场景 1 中的椅子与场景 2 中的桌子的深度层次更加趋于真值.

图 10 为 FlyingThings3D 数据集中的重建结果比较, 由于该数据集通过软件合成, 因此深度真值具有典型的层次结构, 各算法均可以得到较为理想的结果. 从矩形框可以看出, 模型设计类 SFF 在嵌入

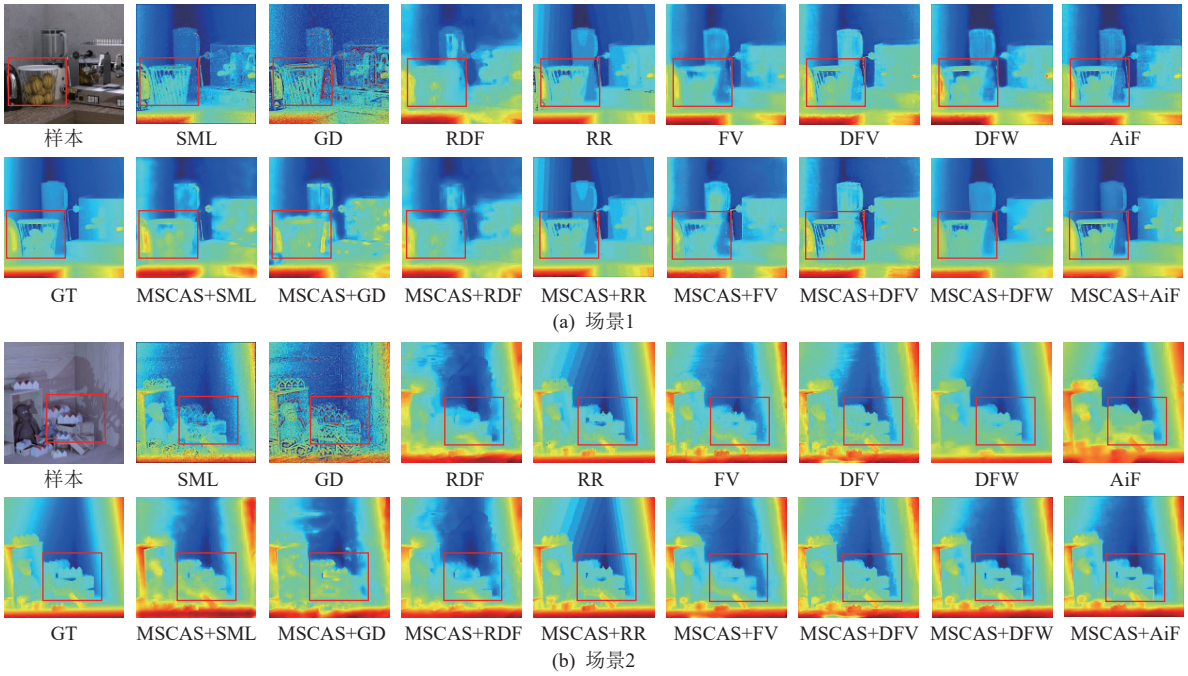


Fig. 8 Results comparison of different algorithms before and after embedding MSCAS framework in LF_4D_PSF dataset

图 8 LF_4D_PSF 数据集中不同算法在嵌入 MSCAS 框架前后的结果对比

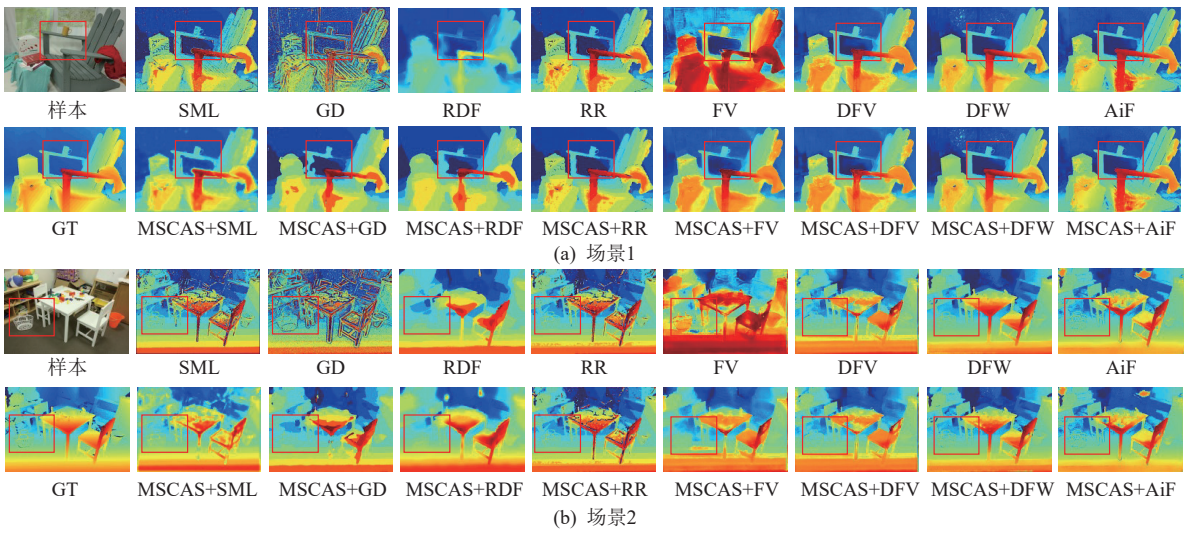


Fig. 9 Results comparison of different algorithms before and after embedding MSCAS framework in Middlebury_FS dataset

图 9 Middlebury_FS 数据集中不同算法在嵌入 MSCAS 框架前后的结果对比

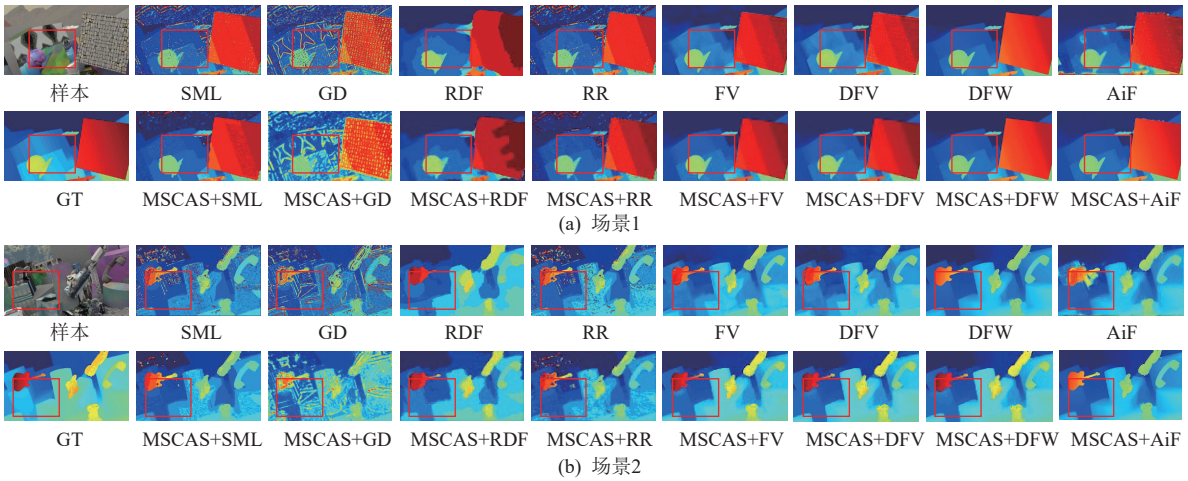


Fig. 10 Results comparison of different algorithms before and after embedding MSCAS framework in FlyingThings3D dataset

图 10 FlyingThings3D 数据集中不同算法在嵌入 MSCAS 框架前后的结果对比

MSCAS 框架后可以去除一些背景噪声，深度学习类 SFF 方法在嵌入 MSCAS 框架后可以得到较为清晰的边缘细节.

综上所述，本文提出的 MSCAS 框架无论是在细节保留还是背景去噪方面均具有一定的优势，而且可以对主流的大多数模型设计类 SFF 方法与深度学习类 SFF 重建方法进行嵌入且实现性能提升，尤其是对模型设计类 SFF 重建方法的性能有较大提升，对深度学习类 SFF 重建方法的边缘细节信息有较好的保留.

4 总 结

为了克服现有多聚焦图像 3 维形貌重建方法在重建过程中的深度信息缺乏和序列间趋势关联信息的缺失，本文提出了一种多尺度代价聚合的多聚焦图像 3 维形貌重建通用框架 MSCAS，该框架引入了非降采样多尺度分解过程实现深度表征信息的倍增，通过多尺度代价聚合实现了尺度间与尺度内关联信息的快速收敛，可对主流的模型设计类 SFF 方法与深度学习类 SFF 方法进行嵌入并实现性能的提升. 在未来的工作中，需要进一步考虑如何根据场景内容实现不同尺度间自适应权重的划分，这将有助于算法的性能进一步提升.

作者贡献声明：闫涛提出算法框架并撰写论文；尚起慧进行实验分析；吴鹏和张江峰提出了论文修改意见；钱宇华指导了论文整体框架设计；陈斌提出了实验设计思路.

参 考 文 献

[1] Nayar S K, Nakagawa Y. Shape from focus[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(8): 824−831

[2] Tao M W, Srinivasan P P, Hadap S, et al. Shape estimation from shading, defocus, and correspondence using light-field angular coherence[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 546−560

[3] Clerc M, Mallat S. The texture gradient equation for recovering shape from texture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 536−549

[4] Lee J Y, Park R H. Complex-valued disparity: Unified depth model of depth from stereo, depth from focus, and depth from defocus based on the light field gradient[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(3): 830−841

[5] Zhan Yu, Guo Xinqing, Lin Haibing, et al. Line assisted light field triangulation and stereo matching[C]//Proc of the 14th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2013: 2792−2799

[6] Li Jianqiao, Lu Mindong, Nian Ze. Continuous depth map reconstruction from light fields[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3257−3265

[7] Yan Tao, Chen Bin, Liu Fengxian, et al. Multi-focus image fusion model for micro 3D reconstruction[J]. Journal of Computer-Aided Design and Computer Graphics, 2017, 29(9): 1613−1623 (in Chinese)
(闫涛, 陈斌, 刘凤娴, 等. 基于多景深融合模型的显微三维重建方法[J]. 计算机辅助设计与图形学学报, 2017, 29(9): 1613−1623)

[8] Yan Tao, Gao Haoxuan, Zhang Jiangfeng, et al. Grouping parallel lightweight real-time microscopic 3D shape reconstruction method[J]. Journal of Software, 2024, 35(4): 1717−1731 (in Chinese)
(闫涛, 高浩轩, 张江峰, 等. 分组并行的实时微观三维形貌重建方法[J]. 软件学报, 2024, 35(4): 1717−1731)

[9] Yan Tao, Qian Yuhua, Li Feijiang, et al. Intelligent microscopic 3D shape reconstruction method based on 3D time-frequency

transformation[J]. SCIENTIA SINICA Informationis, 2023, 53(2): 282−308 (in Chinese)
(闫涛, 钱宇华, 李飞江, 等. 三维时频变换视角的智能微观三维形貌重建方法[J]. 中国科学: 信息科学, 2023, 53(2): 282−308)

[10] Ali U, Mahmood M T. Robust focus volume regularization in shape from focus[J]. IEEE Transactions on Image Processing, 2021, 30: 7215−7227

[11] Yang Fengting, Huang Xiaolei, Zhou Zihan. Deep depth from focus with differential focus volume[C]//Proc of the 35th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 12642−12651

[12] Muhammad M S, Choi T S. Sampling for shape from focus in optical microscopy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3): 564−573

[13] Pertuz S, Puig D, Garcia M A. Analysis of focus measure operators for shape-from-focus[J]. Pattern Recognition, 2013, 46(5): 1415−1432

[14] Ma Z, Kim D, Shin Y G. Shape-from-focus reconstruction using nonlocal matting Laplacian prior followed by MRF based refinement[J]. Pattern Recognition, 2020, 103: 107302

[15] Thelen A, Frey S, Hirsch S, et al. Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation[J]. IEEE Transactions on Image Processing, 2009, 18(1): 151−157

[16] Jeon H G, Surh J, Im S, et al. Ring difference filter for fast and noise robust depth from focus[J]. IEEE Transactions on Image Processing, 2020, 29: 1045−1060

[17] Yan Tao, Wu Peng, Qian Yuhua, et al. Multiscale fusion and aggregation PCNN for 3D shape recovery[J]. Information Sciences, 2020, 536: 277−297

[18] Singh S, Gupta D, Anand R S, et al. Nonsubsampled shearlet based CT and MR medical image fusion using biologically inspired spiking neural network[J]. Biomedical Signal Processing and Control, 2015, 18: 91−101

[19] Minhas R, Mohammed A A, Wu Q M J. Shape from focus using fast discrete curvelet transform[J]. Pattern Recognition, 2011, 44(4): 839−853

[20] Wee C Y, Paramesran R. Measure of image sharpness using eigenvalues[J]. Information Sciences, 2007, 177(12): 2533−2552

[21] Yan Tao, Hu Zhiguo, Qian Yuhua, et al. 3D shape reconstruction from multifocus image fusion using a multidirectional modified Laplacian operator[J]. Pattern Recognition, 2020, 98: 107065

[22] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields[C]//Proc of the 13th Asian Conf on Computer Vision. Berlin: Springer, 2016: 19−34

[23] Won C, Jeon H. Learning depth from focus in the wild[J]. arXiv preprint, arXiv: 2207.09658, 2022

[24] Mayer N, Ilg E, Husser P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]// Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 4040−4048

[25] Maximov M, Galim K, Leal L. Focus on defocus: Bridging the synthetic to real domain gap for depth estimation[C]//Proc of the 35th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 1071−1080

[26] Hazirbas C, Soyer S G, Staab M C, et al. Deep depth from focus[C]//Proc of the 14th Asian Conf on Computer Vision. Berlin: Springer, 2018: 525−541

[27] Wang Ninghsu, Wang Ren, Liu Yulun, et al. Bridging unsupervised and supervised depth from focus via all-in-focus supervision[C]//Proc of the 21st IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 12621−12631

[28] Zhang Jiangfeng, Yan Tao, Wang Keqi, et al. 3D shape reconstruction from multi depth of field images: Datasets and models[J]. Chinese Journal of Computers, 2023, 46(8): 1734−1752 (in Chinese)
(张江峰, 闫涛, 王克琪, 等. 多景深图像聚焦信息的三维形貌重建: 数据集与模型[J]. 计算机学报, 2023, 46(8): 1734−1752)

[29] Liu C, Yuen J, Torralba A. SIFT flow: Dense correspondence across scene and its application[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 978−994

[30] Hosni A, Rheman C, Bleyer M, et al. Fast cost-volume filtering for visual correspondence and beyond[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(2): 504−511

[31] He Kaiming, Sun Jian, Tang Xiaoou. Guided image filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(6): 1397−1409

[32] Yang Qinxiong. A non-local cost aggregation method for stereo matching[C]//Proc of the 25th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 1402−1409

[33] Zhang Kang, Fang Yuqiang, Min Dongbo, et al. Cross-scale cost aggregation for stereo matching[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(5): 965−976

[34] Liang Xinyan, Qian Yuhua, Guo Qian, et al. Multi-granulation fusion-driven method for many-view classification[J]. Journal of Computer Research and Development, 2022, 59(8): 1653−1667 (in Chinese)
(梁新彦, 钱宇华, 郭倩, 等. 多粒度融合驱动的超多视图分类方法[J], 计算机研究与发展, 2022, 59(8): 1653−1667)

[35] Wang Xiaohui, Jia Jia, Cai Lianhong. Expression detail synthesis based on wavelet-based image fusion[J]. Journal of Computer Research and Development, 2013, 50(2): 387−393 (in Chinese)
(王晓慧, 贾珈, 蔡莲红. 基于小波图像融合的表情细节合成[J], 计算机研究与发展, 2013, 50(2): 387−393)

[36] Zhang Lei, Bao P, Wu Xiaolin. Multiscale LMMSE-based image denoising with optimal wavelet selection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(4): 469−481



Yan Tao, born in 1987. PhD, associate professor. Member of CCF. His main research interests include 3D shape reconstruction and machine vision.

闫　涛, 1987 年生. 博士, 副教授. CCF 会员. 主要研究方向为 3 维形貌重建、机器视觉.



Shang Qihui, born in 1998. Master candidate. His main research interests include 3D shape reconstruction and data mining.

尚起慧, 1998年生. 硕士研究生. 主要研究方向为3维形貌重建、数据挖掘.



Wu Peng, born in 1987. PhD, associate professor. Member of CCF. His main research interests include real time operating system and blockchain.

吴　鹏, 1987年生. 博士, 副教授. CCF会员. 主要研究方向为实时操作系统、区块链.



Zhang Jiangfeng, born in 1998. PhD candidate. Student member of CCF. His main research interests include deep learning and 3D shape reconstruction.

张江峰, 1998年生. 博士研究生. CCF学生会员. 主要研究方向为深度学习、3维形貌重建.



Qian Yuhua, born in 1976. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include artificial intelligence and machine learning.

钱宇华, 1976年生. 博士, 教授, 博士生导师. CCF高级会员. 主要研究方向为人工智能、机器学习.



Chen Bin, born in 1970. PhD, professor, PhD supervisor. Member of CCF. His main research interests include artificial intelligence and large language models.

陈　斌, 1970年生. 博士, 教授, 博士生导师. CCF会员. 主要研究方向为人工智能、大语言模型.