

# 改进窗口特征及微分算子的立体匹配算法

李新春<sup>1</sup>, 殷新勇<sup>1\*</sup>, 林森<sup>1,2,3</sup>

<sup>1</sup> 辽宁工程技术大学电子与信息工程学院, 辽宁 葫芦岛 125105;

<sup>2</sup> 中国科学院沈阳自动化研究所机器人学国家重点实验室, 辽宁 沈阳 110016;

<sup>3</sup> 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110016

**摘要** 针对 ELAS(Efficient Large-Scale Stereo Matching)算法视差图条纹明显且具有空洞区域的问题,提出一种匹配窗口特性与微分特性相结合的局部立体匹配算法,增强描述子对点特征信息的描述能力,为待匹配点提供更有区分度的相似性度量。先根据彩色图像的经典自适应算法,从空间上提出适应于灰度图像的窗口描述子,依据图像信号的特点,从像素层面上选择平滑性更小的微分算子。再将匹配窗口与微分算子相结合,获得比只使用两者之一更强的特性信息描述能力。最后进行标准数据集的客观检验与自采集图像的主观评价,结果表明该算法具有较强的鲁棒性和更高的匹配精度,明显改善了原匹配策略视差图中出现条纹及空洞的现象。

**关键词** 机器视觉; 立体匹配; 特征信息描述; 匹配描述子; 水下图像

中图分类号 TN911.73

文献标志码 A

doi: 10.3788/LOP57.081504

## Stereo Matching by Improved Window Characteristics and Differential Operators

Li Xinchun<sup>1</sup>, Yin Xinyong<sup>1\*</sup>, Lin Sen<sup>1,2,3</sup>

<sup>1</sup> School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China;

<sup>2</sup> State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning 110016, China;

<sup>3</sup> Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, Liaoning 110016, China

**Abstract** The parallax pattern obtained from the ELAS (efficient large-scale stereo matching) algorithm contains obvious fringes and void regions. To address this problem, a stereo matching algorithm that combines matching window characteristics with differentials is proposed in this paper. By enhancing the description of the feature information of points, the similarity measure of the points to be matched is provided with a higher degree of discrimination. First, according to the classical adaptive algorithm of color images, a window descriptor adapted to a gray image was proposed spatially. Next, according to the characteristics of an image signal, a less smooth differential operator was selected at the pixel level. Then the proposed matching window was combined with a differential operator to obtain a description ability of feature information stronger than either of the two. Finally, objective evaluation of standard benchmarks and subjective evaluation of self-collected images show that the proposed algorithm is more robust and has higher matching accuracy, and it obviously improves phenomena related to stripes and void regions in the disparity map.

**Key words** machine vision; stereo matching; feature information description; matching descriptor; underwater images

**OCIS codes** 150.0155; 150.1135; 100.2000

收稿日期: 2019-08-29; 修回日期: 2019-09-19; 录用日期: 2019-09-24

基金项目: 国家自然科学基金(61473280,61773367)、国家自然科学基金委员会“共融机器人基础理论与关键技术研究”重大研究计划(91648118)、辽宁省教育厅科学研究一般项目(L2014132)、辽宁省自然科学基金面上项目(2015020100)

\* E-mail: xin-yong.yin@outlook.com

# 1 引言

立体匹配与光流场估计相同,均是三维场景估计的子任务<sup>[1]</sup>。立体匹配的任务是从立体图像对中寻找匹配点,匹配代价计算复杂,长期以来都是计算机视觉的核心问题之一。特别是当病态条件上升后,误匹配率会显著升高<sup>[2]</sup>。尽管立体匹配的研究存在诸多挑战,但在自动驾驶、3D重建和目标检测与识别<sup>[3]</sup>等不同领域的广泛应用,使得该方向自提出以来便成为研究热点。

传统匹配算法根据不同的能量泛函最优化理论,主要分为三类立体匹配技术,即基于局部、基于半全局和基于全局的立体匹配算法<sup>[4-5]</sup>。

随着机器学习在计算机视觉领域的应用,立体匹配的研究方向转变为基于神经网络的匹配方法<sup>[6-8]</sup>,而在文献<sup>[9]</sup>之前仍以传统研究方法为主,从机器学习角度解决立体匹配问题,但未受到广泛关注。同年,传统方法的研究仍占据重要比率,Tanai等<sup>[10]</sup>利用局部共享标签使得GC(Graph Cut)的空间传播得以应用于基于融合的优化,实现了连续的视差估计,但其无法抑制病态区域信息在空间传播;Sinha等<sup>[11]</sup>针对高分辨率图像,利用稀疏特征匹配替代全局搜索,提出局部平面扫描算法生成匹配代价,并用半全局立体匹配(SGM)方法进行全局优化,该算法为保证高效性,只估计了一小部分潜在的视差假设值,因此,当错误假设视差值时,无法约束误匹配点信息在其他预测区域的传播;Zhang等<sup>[12]</sup>根据人类视觉在多尺度上处理立体匹配的仿生概念,提出跨尺度代价聚合可集成各种代价聚合的方法,结合优化中引入的尺度正则化,使得优化更易于求解,但该方法依赖于初始化生成的匹配代价和高斯金字塔的层数,无法针对所有场景提出确定值。卷积神经网络在立体匹配方向不断发展,极大地提高了立体匹配的匹配率,如针对KITTI数据集,文献<sup>[3]</sup>和<sup>[6]</sup>都获得了优异的表现。

虽然机器学习方法极大地提高了立体匹配的匹配率,但目前基于神经网络的立体匹配几乎都属于监督学习<sup>[1,3,6,9]</sup>,即在训练过程中需要真值图或由LiDAR获取的小块视差值,而真值图或LiDAR生成视差块的获取不仅复杂而且耗费财力<sup>[13]</sup>。不仅如此,由训练集切换到测试集或真实场景时,基于机器学习方法的匹配性能会明显下降<sup>[14]</sup>。Pang等<sup>[13]</sup>认为原因有两点:一是训练场景的特征可能与目标场景有很大区别;二是使用不同立体视觉模型在真

实场景中采集的立体图像对,都在不同程度上出现了退化,如噪声腐蚀、光度畸变、矫正缺陷。因而直接将目标场景的立体图像对送入另一个场景中预训练卷积神经网络(CNN),其性能会显著降低。而且训练时需要大量的数据和时间,即使在图形处理器上处理测试集或真实场景中的一对图片也需要1 min以上的时间<sup>[13]</sup>。

相比之下,传统方法面对这些情况时,算法鲁棒性更高。传统匹配过程不依赖真值图,不需要预训练,特征信息的提取只与原图像对有关,可通过几何方法优化匹配代价的计算,降低计算开销<sup>[15]</sup>。为此,本文从传统研究方法出发,着重于像素点的特征描述,以改善特征点数量过少引起视差图出现的明显条纹及空洞(有限远场景出现零值)等问题。先根据彩色图像的经典自适应窗口理论,得到适用于灰度图像的固定大小和形状的窗口描述子,并验证了其他不符合该原理窗口描述子的不适用性。其次针对信号随机性更强的图像,使用Schar滤波器微分图像,从而提升了窗口描述子对待匹配点的特征描述能力。最后使用真值图数据集,客观评估了本文方法与近些年新兴算法的误匹配率;再使用真实场景和水下图像,主观评估场景切换时的算法性能。所提算法在保证高效性的同时降低了误匹配率,明显改善了视差图中易出现条纹和空洞的现象。

## 2 基本原理

源图像特征信息的描述是ELAS(Efficient Large-Scale Stereo Matching)等<sup>[15]</sup>传统算法后续处理步骤的基础,若特征信息处理不当,将直接影响立体匹配算法的性能。从该角度出发,将像素空间分布特征与微分特性结合,同时利用像素的空间信息与微分信息提升匹配描述子对像素点特征信息的描述能力。

### 2.1 点特征信息描述

#### 2.1.1 高斯权重窗口

计算匹配代价需要聚合待匹配点的特征信息,而像素点之间的相互关系可为待匹配点间相似性的度量提供可靠的区分度。传统匹配窗口的设计依赖于彩色空间且着重解决某些纹理特征,未考虑图像信号的特点。但与自适应窗口相比,形状及大小固定的匹配窗口具有复杂度低且方便集成到其他算法的特点。本节从自适应匹配窗口入手,通过推导求得符合灰度图像和固定形状的匹配窗口,再根据滤波算子的大小确定匹配窗口的实际大小。

经典自适应加权法权值  $\omega(p, q)$  由色度空间权值与几何空间权值组成, 权值可表示为

$$\omega(p, q) = kf(\Delta c_{pq})f(\Delta g_{pq}), \quad (1)$$

式中:  $k$  为比例系数;  $\Delta c_{pq}$  为中心像素点  $p$  和其邻域像素点  $q$  在颜色空间上的欧氏距离;  $\Delta g_{pq}$  为  $p$  和  $q$  在坐标空间上的欧氏距离。

凭经验确定的固定大小及形状的匹配窗口, 易出现匹配窗口太小导致聚合信息不足, 窗口太大则引起匹配代价计算开销大的问题。图像可由高斯模型表示<sup>[16-17]</sup>, 将(1)式中  $f(\Delta g_{pq})$  替换成高斯分布权值  $f(\Delta G_{pq})$ , 其中  $\Delta G_{pq}$  为高斯分布权值中的空间临近性。考虑  $du$  方向的微分会保留原图像上的纵向纹理, 而横向纹理弱, 因此对几何权值添加微分方向的约束项  $f(\Delta e_{pq})$  提高相关像素点的权值, 从而提升匹配窗口对特征信息的描述能力, 其中  $\Delta e_{pq}$  为  $p$  的横向几何距离。则改进的自适应公式为

$$\omega(p, q) = kf(\Delta c_{pq})f(\Delta G_{pq})f(\Delta e_{pq}). \quad (2)$$

经典自适应权值算法的色度空间权值由 CIELab 色彩空间定义为

$$f(\Delta c_{pq}) = \exp(-\Delta c_{pq}/\gamma_c), \quad (3)$$

式中: 常数  $\gamma_c$  通常取 7。其中

$$\Delta c_{pq} = \sqrt{(L_p - L_q)^2 + (a_p - a_q)^2 + (b_p - b_q)^2}, \quad (4)$$

$$\mathbf{c}_p = [L_p, a_p, b_p], \quad (5)$$

$$\mathbf{c}_q = [L_q, a_q, b_q], \quad (6)$$

式中:  $\mathbf{c}_p$ 、 $\mathbf{c}_q$  分别为  $p$ 、 $q$  的 CIELab 色彩空间值;  $L_p$ 、 $a_p$ 、 $b_p$  和  $L_q$ 、 $a_q$ 、 $b_q$  分别为  $p$ 、 $q$  在 CIELab 空间中的  $L$  分量、 $a$  分量和  $b$  分量。立体匹配直接处理的数字图像为经变换得到的灰度图, 除单一的灰度值外无其他颜色分量, 灰度空间分配权值比彩色空间适应性更强, 因此  $\mathbf{c}_p$  与  $\mathbf{c}_q$  由三维向量变为一维向量, 即  $\mathbf{c}_p = [I_p]$ 、 $\mathbf{c}_q = [I_q]$ , 则

$$\Delta c_{pq} = \|I_p - I_q\|_1, \quad (7)$$

式中:  $I_p$  与  $I_q$  分别为  $p$  和  $q$  的灰度值。根据图像平滑连续的特点,  $I_p \approx I_q$ , 则  $\Delta c_{pq} \approx 0$ , 此时令  $\Delta c_{pq} \rightarrow 0^+$ , 色度空间权值可推导为

$$f(\Delta c_{pq}) \Big|_{\Delta c_{pq} \rightarrow 0^+} = \lim_{\Delta c_{pq} \rightarrow 0^+} \exp(-\Delta c_{pq}/\gamma_c) = 1. \quad (8)$$

图像信号可由高斯模型表示特性, 即经典自适应算法的几何空间权重满足空间临近性的原则, 引入二维高斯概率密度函数作为几何空间权值分配函数。为简单起见, 使每一维的边缘概率密度函数服从正态分布, 则

$$f(\Delta G_{pq}) = \exp[-(x^2 + y^2)/2] / 2\pi, \quad (9)$$

式中:  $x$ 、 $y$  为以描述子窗口中心为参考点的像素坐标。经内核大小为  $m \times m$  微分后, 只保留单一方向的纹理特征信息, 因此完整保留该方向上的权值是不合理的。为调整微分方向上的权值, 根据待匹配点在  $u$  方向上的几何距离  $\Delta e_{pq}$  对权值重排, 公式为

$$f(\Delta e_{pq}) = \begin{cases} f(\Delta e_{sq}, v_q), & u_p = u_q \\ 1, & u_p \neq u_q \end{cases}, \quad (10)$$

式中:  $u_p$  为点  $p$  沿横向微分方向所在的横向坐标;  $u_q$  为点  $q$  沿横向微分方向所在的横向坐标;  $v_q$  为  $q$  的纵向坐标;  $\Delta e_{sq}$  为待匹配点与邻域内像素点的欧氏距离, 其中  $s = m - p$ 。

已知图像中物体边界的灰度值变化明显, 但灰度值仍保持连续一致特性, 因此匹配窗口越小, 则在物体边界处, (8)式的条件  $\Delta c_{pq} \rightarrow 0^+$  才能成立。结合图像求解梯度时, 使用  $3 \times 3$  大小窗口的 Sobel 算子<sup>[15]</sup>。而 Sobel 算子属于一阶微分, 所以中心像素为  $3 \times 3$  大小的邻域, 实际直接相关的像素区域大小为  $5 \times 5$ , 模型如图 1 所示。

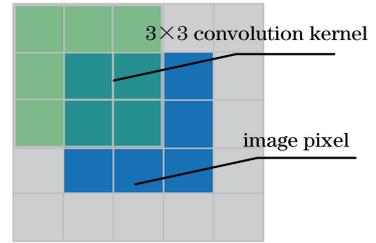


图 1 匹配点直接相关邻域

Fig. 1 Direct relevant neighborhood of matching points

在同时满足计算速度与精度的情况下, 结合 SSE(Streaming SIMD Extensions)指令的匹配窗口描述子共取 16 点, 记  $u$  微分方向为  $du$  取 12 点, 记  $v$  微分方向为  $dv$  取 4 点<sup>[15]</sup>。由此再根据得到的权重值, 可计算得到最大窗口权重  $\omega_{\max}$  及对应的窗口形状, 如图 2 所示。

### 2.1.2 Scharr 滤波器

确定匹配窗口后, 选择与之相适应的微分算子以提高匹配窗口对待匹配点的特征描述至关重要。如文献<sup>[15]</sup>所述, 在统筹考虑算法的时间效率和误匹配率的前提下, 测试了其他稀疏描述子的效果, 最终选择 Sobel 算子作为特征描述子。但测试中并未考虑窗口形状与不同微分算子结合后的效果。

微分算子内核过大, 则破坏(8)式的  $\Delta c_{pq} \rightarrow 0^+$  条件, 无法使用所提描述子窗口; 内核过小, 则对噪声敏感, 无法准确获得特征描述。目前微分算子大

|           |       |       |       |       |       |
|-----------|-------|-------|-------|-------|-------|
|           | 0.029 | 0.013 | 0.022 | 0.013 | 0.029 |
|           | 0.013 | 0.059 | 0.097 | 0.059 | 0.013 |
| $w=0.827$ | 0.097 | 0.022 | 0.159 | 0.022 | 0.097 |
|           | 0.013 | 0.059 | 0.097 | 0.059 | 0.013 |
|           | 0.029 | 0.013 | 0.022 | 0.013 | 0.029 |

图2 窗口形状与权值分配

Fig. 2 Window shape and weight allocation

致分为一阶微分和二阶微分两类,二阶微分如拉普拉斯算子、LoG(Laplacian of the Gaussian)算子和Canny算子等,对图像求取两次导数,因此对灰度变化强烈区域特征描述能力强,丢失一阶微分生成的纹理特征,不利于后续步骤中提取强鲁棒性匹配点。而一阶微分算子如Roberts算子的卷积模板太小,对噪声敏感;而Prewitt算子中心权重值比Sobel算子更小,微分后对图像作过度平滑处理,需要更大的匹配窗口才能获得适当的特征描述。

相对于Sobel算子,Scharr滤波器的中心权重值更大,微分后整幅图像的平滑性更弱。这使得在像素值相似的区域内,微分后结果间的差异更大。例如,具有平滑过渡的凹凸表面且颜色一致的物体,在小区域内像素间的差异很小,使用平滑性更好的Sobel算子,无法扩大微分结果间的差异,导致窗口内的数值总体无明显差异,多点间的匹配代价近似,无法获得最优解致使误匹配。因此,为获得更弱的平滑特性,增强窗口总体的特征描述,使用中心权重值更大的Scharr滤波器可进一步降低误匹配。

源像素经Scharr滤波器卷积内核后,增大像素间的差异。特别是当中心点邻域内像素值差别过小时,Scharr滤波器可提高中心权重值,降低微分算子的平滑性以增大结果间的差异。

## 2.2 匹配代价与视差计算

获得待匹配点的特征信息后,再经曼哈顿距离度量,获得具有强鲁棒性的匹配点,被称为支撑点,记这些点组成的集合为 $\mathbf{S} = \{s_1, s_2, \dots, s_M\}$ ,其中每个支撑点都是一个三维列向量 $s_M = [u_m, v_m, d_m]^T$ ,即图像坐标 $(u_m, v_m) \in \mathbb{N}^2$ 和视差 $d_m \in \mathbb{N}$ 的组合。

假设观测值 $\{\mathbf{o}_n^{(l)}, \mathbf{o}_n^{(r)}\}$ 和支撑点 $\mathbf{S}$ 在视差 $d_n$ 的条件下独立,则先验概率正比于均匀分布和高斯采样的组合形式。 $\mu(\mathbf{S}, \mathbf{o}_n^{(l)})$ 为连接支撑点和观测值的均值函数,可描述为线性分段函数,使用三角剖

分根据支撑点计算所需要的内插视差值。对每个三角形构成的平面定义为

$$\mu_i(\mathbf{o}_n^{(l)}) = a_i u_n + b_i v_n + c_i, \quad (11)$$

式中: $i$ 为像素点 $(u_n, v_n)$ 所属的三角形; $\mathbf{o}_n = [u_n, v_n, f_n]^T$ 为观测值。因此,所提模型 $\mu$ 可粗略表示为支撑点对应的视差之间的线性内插值。

## 2.3 视差值选择与优化

根据2.2节理论获取左右视差图后,依靠最大后验概率计算匹配后的视差图,公式为

$$d_n^* = \operatorname{argmax} P(d_n | \mathbf{o}_n^{(l)}, \mathbf{o}_1^{(r)}, \mathbf{o}_2^{(r)}, \dots, \mathbf{o}_N^{(r)}, \mathbf{S}), \quad (12)$$

式中: $P$ 为概率函数。

在右像平面沿着极线的观测值是结构化的,即给定与 $\mathbf{o}_n^{(l)}$ 相关联的视差后,沿着极线的观测值存在确定性的映射关系,这使其有非零概率。通过建模,沿着极线的所有观测值的分布可获得以下特性:

$$p(\mathbf{o}_1^{(r)}, \mathbf{o}_2^{(r)}, \dots, \mathbf{o}_N^{(r)} | \mathbf{o}_n^{(l)}, d_n) \propto \prod_{i=1}^N p(\mathbf{o}_i^{(r)} | \mathbf{o}_n^{(l)}, d_n). \quad (13)$$

至此,整个算法的流程如图3所示。图3中point feature description处理模块包含所提高斯权重窗口与Scharr滤波器,其中Scharr滤波对应derivation处理流程,高斯权重窗口包含于create descriptor vector处理流程。作为匹配代价的输入源数据,点特征描述信息的优劣直接影响匹配代价计算与匹配代价聚合。为此,将充分考虑空间信息与微分信息联合时对点特征描述的能力,为后续处理步骤提供有效可靠的源数据输入。

## 3 实验结果及分析

为验证高斯权重窗口和Scharr滤波器的有效性,采用Middlebury评估平台<sup>[18]</sup>提供的第三版标准测试数据集,数据集示例见图4。因为新数据集更具挑战性且具有更大的半遮挡区域,评估网站将非遮挡(nonocc)作为默认掩模(实验评估指标下的结果采用相同掩模获得),而客观评价标准使用Middlebury平台示例程序中两项默认评估指标:误匹配率(bad,  $E_{\text{bad}}$ )与绝对值均差(avgErr,  $E_{\text{avg}}$ )。

评价指标bad的含义是在掩模下设置误差限为1 pixel,即当真值图和视差图结果相差1 pixel以上时,判定该点为误匹配点,公式为

$$E_{\text{bad}} = \frac{n}{N} \times 100\%, \quad (14)$$

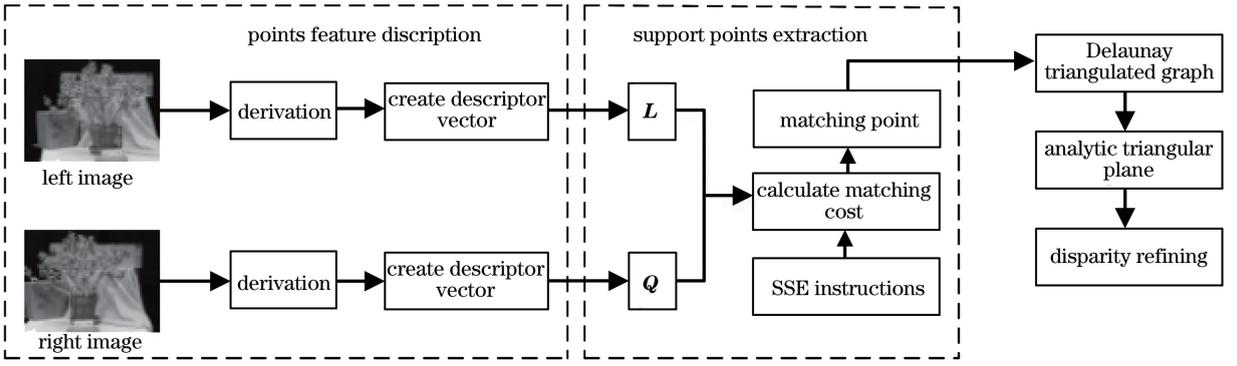


图3 本文匹配主流程图

Fig. 3 Main flow chart of matching in this paper

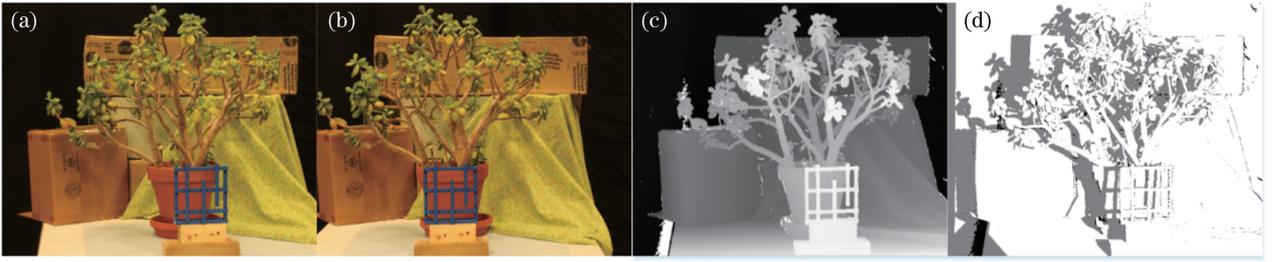


图4 数据集示例图(Jadeplant)。(a)左图像；(b)右图像；(c)真值图；(d)掩模

Fig. 4 Sample datasets diagrams (Jadeplant). (a) Left image; (b) right image; (c) ground truth; (d) mask

式中： $N$  为掩模下有效点的总和； $n$  为超出误差线的匹配点总和。

评价指标  $\text{avgErr}$  的含义是在掩模下计算非遮挡区域真值图的值  $g_i$  与匹配点视差  $d$  (包含误匹配点) 的差的绝对值, 最后统计绝对值的总和与所有有效点的总和的百分比, 即表示得到的视差值与真值图的值的平均总偏差的百分比, 公式为

$$E_{\text{avg}} = \frac{\sum_i^N |d_i - g_{ti}|}{N - I_{\text{inf}}} \times 100\%, \quad (15)$$

式中： $I_{\text{inf}}$  为视差值无限大的无效点的总和。根据(14)式和(15)式可见该项指标与  $\text{bad}$  指标相近时更有意义, 即脱离误匹配率, 单一对比  $\text{avgErr}$  指标并无明显的说服力, (14)式和(15)式间的区别说明该项指标与  $\text{bad}$  指标并不总是一一对应。

误匹配及绝对值均差降低总量计算公式为

$$S_w = \sum_j^M (E_{lj} - P_{rj}), \quad (16)$$

式中： $E_l$  与  $P_r$  分别对应表 1 中的 ELAS 与 Proposed 列； $M$  为对应列的行总数。

实验不仅测试所提方法的有效性与其他高斯权重窗口的不可使用性, 而且设计交叉实验, 验证所提窗口与 Scharr 滤波器结合的有效性。其次, 将所提方

法的匹配性能与典型传统算法作对比; 再对比所提方法与新兴立体匹配算法的误匹配率, 验证所提方法的有效性 with 适用性。最后使用自采集图像验证算法的鲁棒性, 并设计水下图像实验, 验证算法面对水下图像的可行性。编程语言为 C++, 硬件环境为 Intel® Core™ i3 CPU, 内存为 8 GB, 系统为 CentOS7。

### 3.1 高斯权重窗口验证

高斯权重窗口在两种指标下的性能对比结果见表 1。实验结果表明, 在只使用高斯权重窗口的情况下, 对于  $\text{bad}$  指标有 11 组实验结果的性能优于原算法 (共 15 组测试图像); 对于  $\text{avgErr}$  指标有 7 组实验结果的性能优于原算法, 4 组实验结果与原算法相同。

经实验验证, 高斯权重窗口降低误匹配率, 且在更多测试组上取得更优的表现。但值得注意的是, 两个评价指标下匹配性能提高的组别并不对应, 这是因为  $\text{avgErr}$  指标计算的是掩模条件下所有匹配点 (包含误匹配点) 与真值图的值差的绝对值的总和的百分比, 表示的是每一点与真值图的值平均偏离程度, 特别是误匹配点偏离大时会严重影响百分比, 因此该项性能并不总是与  $\text{bad}$  评价指标相对应。实验对比表明高斯权重窗口更具有普适性, 更能适应一般场景切换。

表 1 高斯权重窗口在两种指标下的性能对比

Table 1 Performance comparison of Gaussian weight window under two indices

| Data name   | $E_{bad} / \%$ |          | $E_{avg} / \%$ |          | Data name   |
|-------------|----------------|----------|----------------|----------|-------------|
|             | ELAS           | Proposed | ELAS           | Proposed |             |
| Adirondack  | 12.070         | 11.890   | 0.880          | 0.877    | Adirondack  |
| ArtL        | 19.800         | 19.270   | 1.760          | 1.630    | ArtL        |
| Jadeplant   | 34.680         | 34.660   | 1.170          | 1.170    | MotorcycleE |
| MotorcycleE | 13.450         | 13.210   | 2.210          | 2.080    | PianoL      |
| PianoL      | 37.520         | 36.530   | 1.960          | 1.930    | Pipes       |
| Pipes       | 15.300         | 14.760   | 1.900          | 1.800    | layroom     |
| Playroom    | 26.680         | 26.040   | 4.550          | 4.270    | Playtable   |
| Playtable   | 26.380         | 26.290   | 0.930          | 0.920    | PlaytableP  |
| PlaytableP  | 14.840         | 14.740   | 0.890          | 0.890    | Recycle     |
| Recycle     | 17.850         | 16.850   | 0.910          | 0.910    | Teddy       |
| Vintage     | 33.690         | 33.660   | 2.430          | 2.430    | Vintage     |

为进一步验证高斯权重窗口的有效性,在图 5 所示结果中,不仅通过实验验证其他权重对应窗口形状的不适用性,且验证窗口不符合  $5 \times 5$  大小的不适用性(对应窗口形状见图 6)。通过分析多个窗口匹配结果,验证高斯权重窗口的有效性,即高斯权重窗口可更有效地描述匹配点特征信息。其他窗口描述匹配点特征信息时,无法捕获邻域范围内的信息,造成特征信息丢失,直接表现为窗口内像素值总体上类似,导致在计算匹配代价时无法区分类似匹配点,最终出现误匹配。

分析图 5 结果时,注意到这样一个事实:虽然不符合高斯权重窗口的误匹配率高,在测试集上大部分实验组的性能不佳,但在个别图像上一些形状的窗口性能取得更优的表现,如 D 型窗口,对 Jadeplant 图像的误匹配率低于其他窗口,但在其他图像上的误匹配率更高。分析原因,所提窗口可应

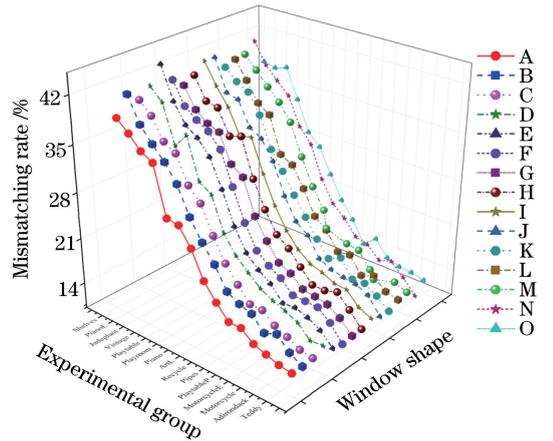


图 5 不同窗口形状误匹配率

Fig. 5 Mismatching rate of different window shapes  
对更多的场景类型,适应不同病态条件,但当病态条件超过其适应能力,影响图像信号的高斯模型时,特殊类型窗口可更好地描述特征信息。

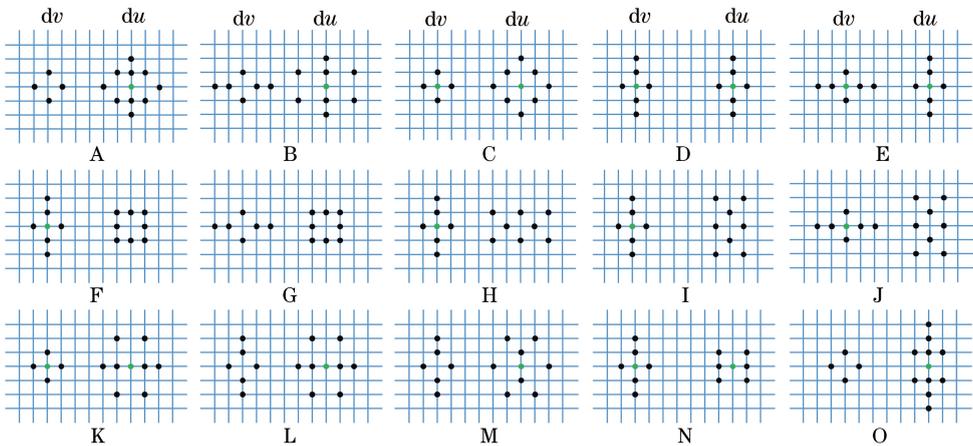


图 6 不同类型的匹配窗口

Fig. 6 Different types of matching windows

由图 6 可以看到, A 型窗口是符合实验原理的窗口, 中心点表示取点两次(该点权重是次最大权重值的近 2 倍)。实验中修改  $dv$  和  $du$  的点数和形状(总点数保持为 16), 证明其他窗口的无效性。类似于图中标号为 O 的窗口类型, 其与实验使用的窗口(A 型)只是  $du$  的形状不同, 根据图 5 的结果该窗口性能并不如 A 型窗口, 实际上类似于 O 型窗口。还测试了其他形状, 总体性能并不如所提窗口类型, 但限于篇幅, 文中不一一列出。

### 3.2 Scharr 滤波器验证

Scharr 滤波器通过增大中心权重值, 降低像素点间的平滑性, 使得微分后的数据项更独立。当邻域内像素值差别很小时, 可提高邻域内匹配点的特征描述

能力, 为匹配代价的计算提供更有力的支撑。Scharr 滤波器在两种指标下的性能对比结果见表 2, 对于 bad 指标有 9 组实验结果的性能优于原算法; 对于 avgErr 指标有 5 组实验结果的性能优于原算法。

Scharr 滤波器在 bad 评价指标下取得优势, 而在 avgErr 评价指标下表现出相对劣势, 实际上在 PianoL 和 Pipes 实验组上 Scharr 与原算法的误匹配率相差甚小(相差 0.01)。分析原因, Scharr 滤波器在微分上增加中心权重值, 适当缩小像素间的平滑性, 但改进并未改变 Sobel 算子微分的性质, 而是提升 Sobel 算子在数字图像上的性能。综合两个评价指标, Scharr 滤波器总体上取得优于原算法的性能, 因此这种提升是有效的。

表 2 Scharr 滤波器在两种指标下的性能对比

Table 2 Performance comparison of Scharr filter under two indices

| Data name   | $E_{bad}/\%$ |          | $E_{avg}/\%$ |          | Data name   |
|-------------|--------------|----------|--------------|----------|-------------|
|             | ELAS         | Proposed | ELAS         | Proposed |             |
| ArtL        | 19.80        | 19.73    | 0.88         | 0.86     | Adirondack  |
| Jadeplant   | 34.68        | 29.19    | 8.85         | 6.33     | Jadeplant   |
| MotorcycleE | 13.45        | 13.09    | 1.17         | 1.13     | MotorcycleE |
| Piano       | 22.04        | 21.92    | 2.21         | 2.22     | PianoL      |
| PianoL      | 37.52        | 37.50    | 1.96         | 1.97     | Pipes       |
| Pipes       | 15.30        | 15.19    | 0.93         | 0.92     | PlaytableP  |
| Playroom    | 26.68        | 25.63    | 0.91         | 0.89     | Teddy       |
| PlaytableP  | 14.84        | 14.79    | —            | —        | —           |
| Recycle     | 17.85        | 17.24    | —            | —        | —           |

### 3.3 交叉验证

在 3.2 节实验基础上, 交叉验证的目的之一是验证高斯权重窗口与 Scharr 滤波器结合后的性能(未经特殊说明, “所提方法”均指两者结合的方法), 通过增强窗口特性与微分特性进一步提升匹配窗口

的特征描述特性。所提方法在两种指标下的匹配性能对比结果见表 3。实验结果表明, 所提方法在 bad 指标下有 10 组实验结果的性能优于原算法, 1 组实验结果与原算法表现性能相同; 对于 avgErr 指标有 11 组实验结果的性能优于原算法。

表 3 所提方法在两种指标下的匹配性能对比

Table 3 Performance comparison of proposed algorithm under two indices

| Data name   | $E_{bad}/\%$ |          | $E_{avg}/\%$ |          | Data name   |
|-------------|--------------|----------|--------------|----------|-------------|
|             | ELAS         | Proposed | ELAS         | Proposed |             |
| Adirondack  | 12.07        | 11.81    | 0.88         | 0.83     | Adirondack  |
| ArtL        | 19.80        | 19.55    | 1.76         | 1.74     | ArtL        |
| Jadeplant   | 34.68        | 27.85    | 8.95         | 5.99     | Jadeplant   |
| MotorcycleE | 13.45        | 13.30    | 1.17         | 1.15     | Motorcycle  |
| Piano       | 22.04        | 21.85    | 1.17         | 1.16     | MotorcycleE |
| PianoL      | 37.52        | 36.03    | 2.21         | 1.94     | PianoL      |
| Pipes       | 15.30        | 15.30    | 1.96         | 1.94     | Pipes       |
| Playroom    | 26.68        | 25.63    | 0.93         | 0.92     | PlaytableP  |
| Recycle     | 17.85        | 17.12    | 1.97         | 1.92     | Shelves     |
| Teddy       | 11.42        | 10.88    | 0.91         | 0.83     | Teddy       |
| Vintage     | 33.69        | 32.33    | 2.43         | 2.39     | Vintage     |

经实验验证,所提方法不仅在 bad 指标下的性能优于单独使用高斯权重窗口和 Scharrr 滤波器(误匹配率总降低量分别为 12.85 个百分点、4.36 个百分点和 7.88 个百分点),在 avgErr 指标下的性能同样优于单独使用高斯权重窗口和 Scharrr 滤波器(绝对值均差总降低量分别为 3.530 个百分点、0.683 个百分点和 2.710 个百分点)。图 7 给出了实验效果提升示例,方框标注所提方法在视差图上的提升效果。在 Vintage 与 Jadeplant 场景中,所提方法的视差图条纹明显减少,纹理性更好;在 PianoL 场景中,所提方法产生的零值区域更少。说明所提算法达到降低误匹配率的效果,并改善了视差图中的条纹及

空洞现象。

所提方法在单对图像上的运行时间最多降低 0.03 s,总运行时间平均降低 0.0033 s。可见所提方法几乎不影响算法的运行时间,即所提方法并未试图牺牲匹配效率来提高匹配率。

相同分辨率下,图 8(评价指标 bad)和图 9(评价指标 avgErr)将实验结果与新兴算法和传统算法作对比,其中 GwcNet (Group-wise Correlation Stereo Network)<sup>[19]</sup> 和 GA-Net (Guided Aggregation Network)<sup>[20]</sup> 均为基于 CNN 的新兴立体匹配算法,而传统匹配算法则使用局部立体匹配算法 Census 与半全局立体匹配(SGM)算法。

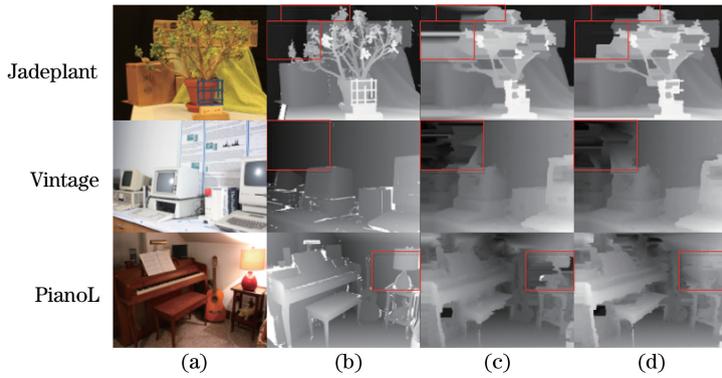


图 7 所提方法与原方法效果提升示例。(a)左图像;(b)真值图;(c)ELAS;(d)本文算法

Fig. 7 Example of the effect improvement of proposed algorithm and the original method. (a) Left images; (b) ground truth images; (c) ELAS; (d) proposed algorithm

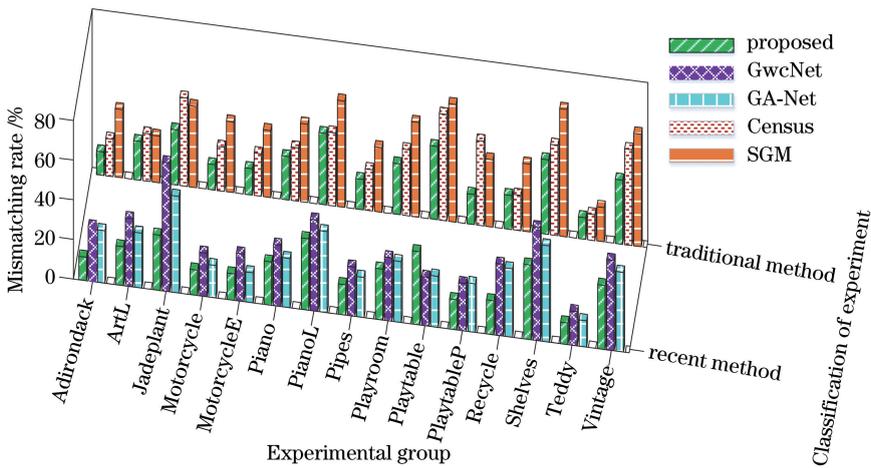


图 8 bad 指标对比图

Fig. 8 bad index comparison

综合分析实验结果,可见 GA-Net 的匹配性能优于 GwcNet 算法,因为 GwcNet 在匹配过程中只使用卷积神经网络进行分组相关的代价聚合计算,而 GA-Net 在匹配过程中,利用卷积神经网络引导匹配代价的聚合,使得匹配代价在计算过程中尽可能避免病态区域信息在神经网络中传播;然而所提

方法优于这两种新兴方法(除 Playtable 图像对外),分析认为在测试时使用预先训练好的模型,而测试数据非 KITTI 数据集,如第 1 节所述,当切换场景时,由于缺少训练场景的特征信息,基于机器学习的立体匹配算法的匹配率明显下降,然而所提方法不依赖于训练模型,取得更优的性能;其次所提方法在

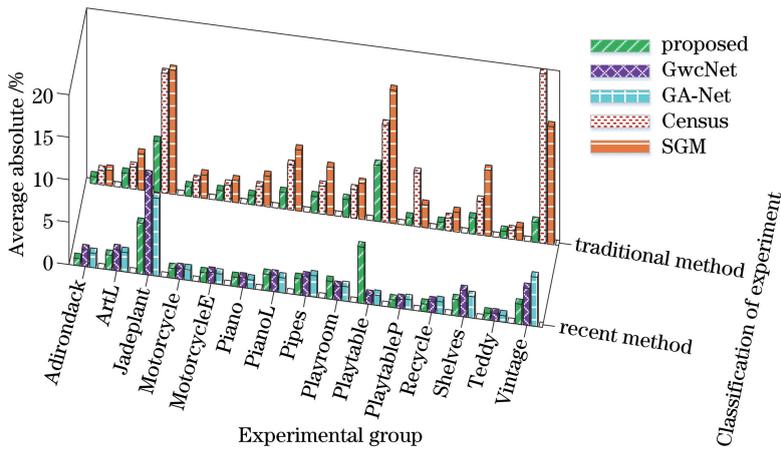


图 9 avgErr 指标对比图  
Fig. 9 avgErr index comparison chart

PlaytableP 的图像对上的匹配性能要高于 Playtable, 原因在于前者为完美校正的图像对, 而所提方法是传统匹配方法, 匹配点的搜索被限制在对准行上, 当行对准失效后将严重影响匹配性能; 而 GwcNet 和 GA-Net 均使用互信息, 对于行对准的需求更低, 在 Playtable 图像对上的匹配性能也更好; 另有四组图像对 (即 Pinao、PianoL、Playroom 和 Teddy) 的 avgErr 值略低于两种新兴算法, 分析原因在于由互信息得到的视差图区域一致性更强, 导致误匹配的视差值更接近附近匹配点的视差值, 因而部分图像对的该项指标略高于所提方法, 但在整体上, 所提方法的该项指标取得绝对优势 (10/15 的图像对该项性能更优), 另外这四组图像对的 bad 指标明显低于所提方法, 在 avgErr 值近似时 bad 指标低更有说服力。综上所述, 当切换测试场景时, 基于机器学习的立体匹配性能低于所提方法, 所提方法并不依赖于某一特定场景的特征信息, 因而可面对各类视觉场景, 更符合机器视觉发展的一般要求。

所提方法同样优于测试的两个传统算法的匹配性能, 这是因为实验使用的数据集更具有挑战性, 即病态条件严重, 而 SGM 算法聚合特定方向上的全局信息, 导致病态区域的信息可不同程度地扩散到整幅图像上, 影响 SGM 算法的匹配性能; Census 变换和所提方法在计算匹配代价时都使用大小一定的匹配窗口聚合信息, 抑制病态信息的传播。但 Census 变换只保留窗口内像素之间的关系, 造成源图像灰度信息的丢失, 这在灰度值近似区域易造成误匹配, 而所提方法避免了这一现象, 获得更高的匹配率。

在主观上, 视差图对比示例见图 10, 方框内标识所提方法明显提升的效果图。其中 GA-Net 视差图与真值图相比总体亮度太暗, 原因是整体视差值偏小; 而 GwcNet 视差图的纹理不明显, 这表明其无法区分有凹凸形状的物体, 视差图无法体现物体的细致纹理。对于 Census 视差图, 遮挡区域出现大面积条纹, 视差图纹理不清晰且有间断区域; 而 SGM 视差图在前景物体上出现零值区域, 物体形状

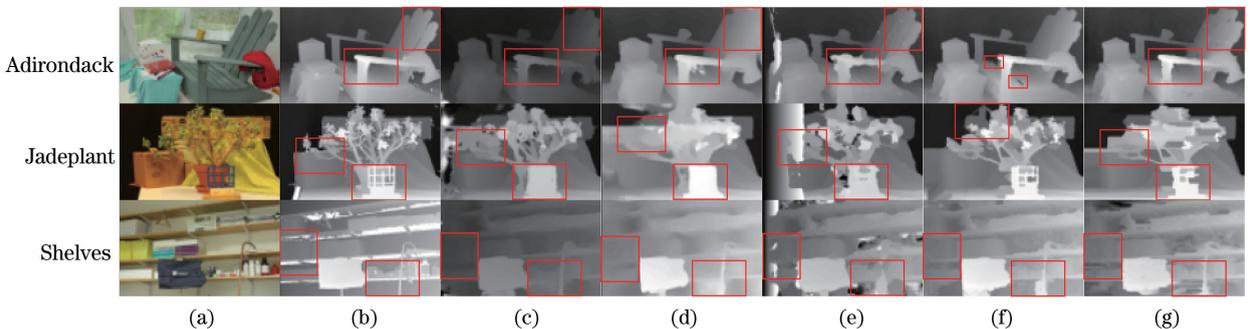


图 10 提升效果对比示例图。(a)左图像;(b)真值图;(c)GA-Net;(d)GwcNet;(e)Census;(f)SGM;(g)本文算法

Fig. 10 Sample graphs of effectiveness improvement. (a) Left images; (b) ground truth; (c) GA-Net; (d) GwcNet; (e) Census; (f) SGM; (g) proposed algorithm

出现变形(如 Jadeplant 视差图),且 Shelves 的视差图纹理不清晰。

### 3.4 真实场景与水下图像测试

为进一步验证所提方法的鲁棒性,实验测试自采集场景图,由于缺少真值图,实验数据无法进行客观对比,只能主观上对比算法间的优劣,结果见图 11。根据双目视觉深度公式  $Z = fT/d$  ( $f$  为相机焦距,  $T$  为双目摄像头的水平距离)和真值图特点,获得几点主观评价标准:1)纹理清晰,视差连续;2)室内场景无视差值为零的区域;3)前景物体的视差值比背景大。

对于图 11 中两幅陆地图像(Air\_1 和 Air\_2),传统匹配方法 Census 和 SGM 生成的视差图均明显出现大量零值区域,在源图像前景与背景单一的情况下,其视差图不连续,并且图像纹理发生严重扭

曲和变形(如 Air\_2 中画板的边沿)。而所提方法生成的视差图纹理明显且连续,前景与背景区分度更大;算法 GwcNet 与 GA-Net 也表现出这些特点,但是如 Air\_1 对应的各算法视差图所示,书本属于前景物体,在 GwcNet 与 GA-Net 的视差图中几乎与背景融为一体,而所提方法生成的视差图中却有更明显的纹理边界,且书本的视差值比周围数值更大。对于 Air\_2 对应的视差图,GwcNet 与所提方法生成的视差值整体近似,而 GA-Net 生成的视差值整体偏小,根据实际场景的拍摄情况,GwcNet 与所提方法的视差图更符合实际情况。但是在遮挡区域(在该测试图中属于背景),GwcNet 的视差值与柱子几乎相同,而实际上视差值应该更小;这一点所提方法虽然产生小部分零值区域,但在遮挡区域的视差值低于柱子的视差值,更符合实际场景的情况。

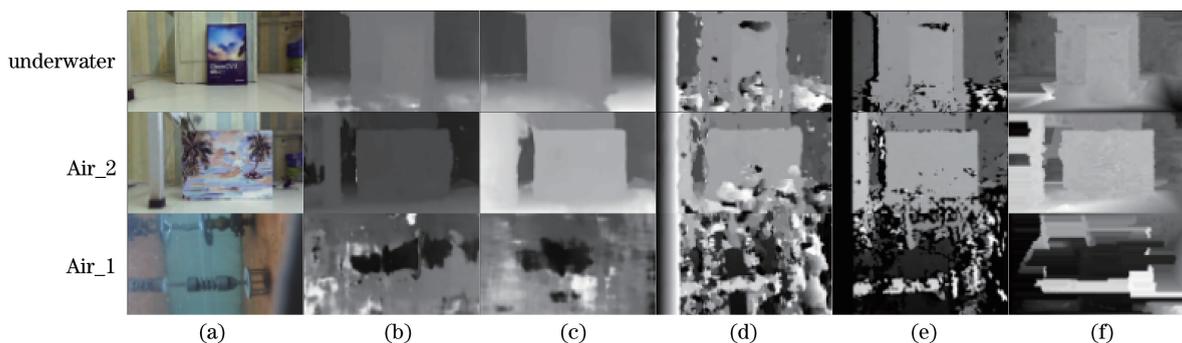


图 11 自采集图像。(a)左图像;(b) GA-Net;(c) GwcNet;(d) Census;(e) SGM;(f)本文方法

Fig. 11 Self-collecting images. (a) Left images; (b) GA-Net; (c) GwcNet; (d) Census;

(e) SGM; (f) proposed algorithm

水下图像由于受到吸收、散射和颜色失真的影响,各算法生成的视差图很不理想。实验测试中所使用的传统算法和机器学习算法生成的视差图,都表现出很低的匹配性能。如 Census 和 SGM 算法的视差图存在大量零值区域和视差不连续区域,无法区分前景与背景;而两个机器学习算法生成的视差图,虽然有更少的零值区域,但前景物体无大致轮廓,前景与背景间也无明显层次感。在所提方法生成的视差图中,虽然前景物体也无清晰轮廓,但具有两个前景物体的大致轮廓,与背景间有明显层次感,而且视差图连续性更差,并与两个传统算法相比具有更少的零值区域。因而,结合水下图像的匹配性能,所提方法的鲁棒性更高,能更好地应对复杂场景,并且不受场景切换的影响。

## 4 结 论

提出了一种改进窗口特征及微分算子的立体匹

配算法,用于增强待匹配点的特征描述信息,进而降低误匹配率,达到改善视差图条纹与空洞现象的目的。同时利用待匹配点与邻近像素点的几何信息,以及像素点微分信息,增强了邻近像素点间匹配代价的区分度。实验结果表明,所生成视差图的条纹与空洞区域得到明显改善,与原算法相比误匹配率降低总和为 12.85 个百分点(10/15 组的优势性能),avgErr 指标降低总和为 3.53 个百分点(11/15 组的优势性能)。此外,所提算法作为一种增强待匹配点特征信息的算法,易与其他立体匹配算法结合,从而改善原始算法生成的视差图效果,降低误匹配率。

## 参 考 文 献

- [1] Mayer N, Ilg E, Hausser 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 (CVPR), June 27-30, 2016, LasVegas.

- NV, USA. New York: IEEE, 2016: 4040-4048.
- [2] Kendall A, Martirosyan H, Dasgupta S, et al. End-to-end learning of geometry and context for deep stereo regression [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 66-75.
- [3] 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, 2018: 5410-5418.
- [4] Li D H, Shen H Y, Yu X, et al. Binocular ranging method using stereo matching based on improved Census transform [J]. *Laser & Optoelectronics Progress*, 2019, 56(11): 111503.  
李大华, 沈洪宇, 于晓, 等. 一种改进 Census 变换的双目匹配测距方法[J]. *激光与光电子学进展*, 2019, 56(11): 111503.
- [5] Wang Q, Piao Y. Depth image acquisition technology based on improved genetic algorithm [J]. *Laser & Optoelectronics Progress*, 2018, 55(2): 021003.  
王琦, 朴燕. 基于改进遗传算法的深度图像获取技术[J]. *激光与光电子学进展*, 2018, 55(2): 021003.
- [6] Zhang Y, Khamis S, Rhemann C, et al. Active Stereo Net: end-to-end self-supervised learning for active stereo systems [C] // Proceedings of the European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. New York: IEEE, 2018: 784-801.
- [7] Žbontar J, LeCun Y. Computing the stereo matching cost with a convolutional neural network [C] // Proceedings of the IEEE conference on computer vision and pattern recognition, June 7-12, 2015, Boston, USA. New York: IEEE, 2015: 1592-1599.
- [8] Xiao J S, Tian H, Zou W T, et al. Stereo matching based on convolutional neural network [J]. *Acta Optica Sinica*, 2018, 38(8): 0815017.  
肖进胜, 田红, 邹文涛, 等. 基于深度卷积神经网络的双目立体视觉匹配算法[J]. *光学学报*, 2018, 38(8): 0815017.
- [9] Spyropoulos A, Komodakis N, Mordohai P. Learning to detect ground control points for improving the accuracy of stereo matching [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, Ohio. New York: IEEE, 2014: 1621-1628.
- [10] Tanaii T, Matsushita Y, Naemura T. Graph cut based continuous stereo matching using locally shared labels[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, Ohio. New York: IEEE, 2014: 1613-1620.
- [11] Sinha S N, Scharstein D, Szeliski R. Efficient high-resolution stereo matching using local plane sweeps [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, Ohio. New York: IEEE, 2014: 1582-1589.
- [12] Zhang K, Fang Y, Min D, et al. Cross-scale cost aggregation for stereo matching [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, Ohio. New York: IEEE, 2014: 1590-1597.
- [13] Pang J, Sun W, Yang C, et al. Zoom and learn: generalizing deep stereo matching to novel domains [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 18-22, 2018, Salt Lake City, Utah. New York: IEEE, 2018: 2070-2079.
- [14] Tonioni A, Tosi F, Poggi M, et al. Real-time self-adaptive deep stereo [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 16-20, 2019, Long Beach, CA. New York: IEEE, 2019: 195-204.
- [15] Geiger A, Roser M, Urtasun R. Efficient large-scale stereo matching [M] // Kimmel R, Klette R, Sugimoto A. *Computer vision-ACCV 2010. Lecture notes in computer science*. Heidelberg: Springer, 2010, 6492: 25-38.
- [16] Wu P L, Li Y N, Yang F, et al. A CLM-based method of indoor affordance areas classification for service robots [J]. *Robot*, 2018, 40(2): 188-194.  
吴培良, 李亚南, 杨芳, 等. 一种基于 CLM 的服务机器人室内功能区分方法[J]. *机器人*, 2018, 40(2): 188-194.
- [17] Wu S, Xu J Z, Zhang Y X, et al. Reliability evaluation method and application for light-stripe-center extraction [J]. *Acta Optica Sinica*, 2011, 31(11): 1115001.  
王顺, 徐静珠, 张益昕, 等. 结构光光条中心点信度评价方法与应用[J]. *光学学报*, 2011, 31(11): 1115001.
- [18] Scharstein D, Szeliski R, Hirschmüller H. Middlebury stereo vision page [EB/OL]. (2014-10-

- 15) [2019-08-26]. <http://vision.middlebury.edu/stereo/>.
- [19] Guo X, Yang K, Yang W, et al. Group-wise correlation stereo network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 16-20, 2019, Long Beach, CA. New York: IEEE, 2019: 3273-3282.
- [20] Zhang F, Prisacariu V, Yang R, et al. GA-Net: guided aggregation net for end-to-end stereo matching [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 16-20, 2019, Long Beach, CA. New York: IEEE, 2019: 185-194.