

改进窗口特征及微分算子的立体匹配算法

李新春¹, 殷新勇^{1*}, 林森^{1,2,3}

¹ 辽宁工程技术大学电子与信息工程学院, 辽宁 葫芦岛 125105;

² 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;

³ 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110016

对象是谁? ELAS算法; 对象是什么? 立体匹配算法
对象怎么了? 存在问题

针对这个问题, 提出什么算法(近似论文题目), 实现什么效果

从操作层面说下算法的流程和优势

摘要 ELAS 算法是一种性能优良的典型传统立体匹配算法, 但该算法视差图条纹明显且具有空洞区域。针对这一问题, 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, 通过增强描述子对点特征信息的描述, 为待匹配点提供更有区分度的相似性度量。首先根据彩色图像的经典自适应算法, 从空间上提出了适应于灰度图像的窗口描述子, 该窗口在待匹配点的邻域内具有固定大小和形状的特征, 避免在匹配过程中反复计算匹配窗口。其次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。然后将提出的匹配窗口与微分算子相结合, 获得了比只使用两者之一更强的特性信息描述能力。最后通过标准数据集客观检验与自采集图像主观评价, 表明算法更为稳健, 而且具有更高的匹配精度, 明显改善了原匹配策略视差图中出现的条纹及空洞的问题。

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 **文献标识码** A

用事实体现本文的算法的有效性

Stereo Matching by Improved Window Characteristics and Differential Operators

Li Xinchun¹, Yin Xinyong^{1*}, Lin Sen^{1,2,3}

¹ School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China;

² State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences., Shenyang, Liaoning, 110016, China;

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

Abstract ELAS algorithm is a typical traditional matching algorithm with good performance, but the disparity maps of the algorithm has obvious stripes and void regions. To resolve these issues, a stereo matching algorithm is proposed in this paper, which combines the characteristics of matching window with the differentials. By enhancing the descriptor's feature information of the points, the discrimination of similarity measure is provided for the matching points. Firstly, according to the classical adaptive algorithm of color images, a window descriptor adapted to the gray image is proposed spatially. The window feature has a fixed size and shape in the field of the matching points, which avoids repeatedly calculating the matching window in the matching process. Next, according to the characteristics of image signal, the less

收稿日期: 年-月-日; 修回日期: 年-月-日; 录用日期: 年-月-日

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;

smoothness differential operator is selected at the pixel level. Then the proposed matching window is combined with the differential operator to obtain a stronger description ability of feature information than either of the two. Lastly, through the objective evaluation of standard benchmark and subjective evaluation of self-collected images show that the proposed algorithm is more robust and has higher matching accuracy, and obviously improves the stripes issues and minimizes 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

1 引 言

立体匹配与光流场估计一样，是三维场景估计的子任务^[1]。立体匹配的任务是从立体图像中寻找匹配点。其匹配代价计算复杂，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^[2]。尽管立体匹配的研究存在诸多挑战，但其在自动驾驶、3D 重建和目标检测与识别^[3]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的热点。

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：基于局部、基于半全局和基于全局的立体匹配算法^[4-5]。

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也转变为基于神经网络的匹配方法^[6-8]，而在文献[9]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并未受到广泛关注，例如同年传统方法的研究仍然占据重要比重：Taniai 和 Matsushita 等人^[10]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，但是其无法抑制病态区域信息在空间传播信息；Sinha 和 Daniel 等人^[11]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以生成匹配代价，并用半全局立体匹配方法(Semi-global matching, SGM)进行全局优化，由于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，当错误假设视差值时，无法约束误匹配点信息在其它预测区域的传播；Zhang 和 Fang 等人^[12]根据人类视觉在多尺度上处理立体匹配的仿生概念，提出的跨尺度代价聚合可以集成各种代价聚合方法，结合优化中引入的尺度正则化，使得优化更易于求解，但是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，极大提高了立体匹配的匹配率，如面对 KITTI 数据集，文献[3]和[6]都获得了优异的表现。

虽然机器学习方法使匹配率极大提高,但是目前基于神经网络的立体匹配几乎都属于监督学习^[1,3,6,9],即在训练过程中需要 ground truth 图或者由 LiDAR 获取的小块视差值,而 ground truth 图或 LiDAR 生成的视差块的获取不仅复杂而且耗费财力^[13]。不仅如此,当由训练集切换到测试集或真实场景中时,基于机器学习的方法匹配性能会明显下降^[14]。Pang 和 Sun 等人^[13]认为原因有两点:一是训练场景中可能与目标场景中的特征有很大区别;一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,例如: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在另一个场景中预训练的 CNN (convolutional neural networks) 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 GPU 上处理测试集或真实场景中的一对图片也需要一分多钟^[13]。

相比之下,传统方法面对这些情况时,算法鲁棒性更高。即传统匹配过程不依赖 ground truth 图,不需要预训练,特征信息的提取只与原图像对有关,能通过几何方法优化匹配代价的计算,以降低计算开销^[15]。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。首先,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,并验证了其它不符合该原理的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 Scharr 滤波器微分图像,从而提高了窗口描述子对待匹配点的特征描述。最后使用具有 ground truth 图的数据集,客观评估了本文方法与近年新颖算法的误匹配率;再使用真实场景和水下图像,主观评估了场景切换时的算法性能。提出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的条纹和空洞问题。

2 基本原理

对源图像特征信息的描述是 ELAS 等^[15]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,本文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高匹配描述子对像素点特征信息的描述。

2.1 点特征信息描述

2.1.1 高斯权重窗口

匹配代价的计算需要聚合待匹配点的特征信息,而像素点之间的相互关系能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注重解决某些纹

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它算法的特点。本节从自适应匹配窗口入手，通过推导求得符合灰度图像的固定大小和形状的匹配窗口，再根据滤波算子的大小确定窗口描述子的实际大小。

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

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

公式 (1) 中 k 为比例系数， Δc_{pq} 表示待匹配点 p 和其邻域像素点 q 在颜色空间上的欧氏距离， Δg_{pq} 表示 p 和 q 在坐标空间上的欧氏距离。

对于凭经验确定的固定大小及形状的匹配窗口，容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；再结合图像可由高斯模型表示的特点^[16-17]。将 $f(\Delta g_{pq})$ 项替换成 $f(\Delta G_{pq})$ 高斯分布权值，其中 ΔG_{pq} 表示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；因此对几何权重值添加微分方向的约束项 $f(\Delta e_{pq})$ 提高相关像素点的权重值，从而提升匹配窗口对特征信息的描述能力，其中 Δe_{pq} 表示中心像素 p 横向几何距离。则改进的自适应公式为：

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

经典自适应权值算法的色度空间权值由 CIELab 色彩空间定义为下式：

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

公式 (3) 中 c_p 、 c_q 分别表示 p 、 q 点的 CIELab 色彩空间值，常数 γ_c 通常取 7； Δc_{pq} 表示为待匹配点与邻域内像素点的欧氏距离，其中：

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

$$c_p = [L_p, a_p, b_p], \quad (5)$$

$$c_q = [L_q, a_q, b_q], \quad (6)$$

公式 (4) 中 L_p 、 a_p 、 b_p 和 L_q 、 a_q 、 b_q 分别是像素点 p 、 q 在 CIELab 空间中的 L 分量、 a 分量和 b 分量。而立体匹配直接处理的数字图像信号为经过变换得到的灰度图，除单一的灰度值外无其它颜色分量，因而根据灰度空间分配权重值比彩色空间适应性更强，因此 c_p 与

c_q 由三维向量变为一维向量 $c_p = [I_p]$ 、 $c_q = [I_q]$ ，则

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

公式 (7) 中 I_p 与 I_q 分别表示待匹配点 p 和其邻域像素点 q 的灰度值。依据图像平滑连续的特点，待匹配点邻域内像素值近似 $I_p \approx I_q$ ，则 $\Delta c_{pq} \approx 0$ ，此时令 $\Delta c_{pq} \rightarrow 0^+$ 色度空间权值可推导为下式：

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

根据图像信号能由高斯模型表示的特性，及经典自适应算法的几何空间权重满足空间临近性的原则，将二维高斯概率密度函数引入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：

$$f(\Delta G_{pq}) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}, \quad (9)$$

公式 (9) 中 x 、 y 表示以描述子窗口中心为参考点的像素坐标。经过内核大小 $m \times m$ 的微分后，只保留了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 Δe_{pq} 对权值重排，公式如下：

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

公式 (10) 中 u_p 表示待匹配点沿横向微分方向所在的横向坐标， v_q 表示待匹配点的邻域像素点的纵向坐标，其中 $s = m - p$ 。

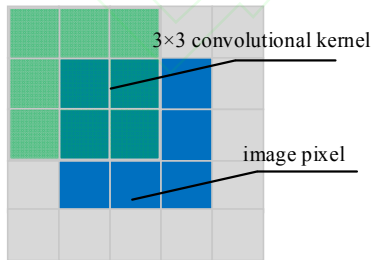


图 1 匹配点直接相关邻域

Fig. 1 Direct relevant neighborhood of matching points

	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 allocate

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，则在物体边界处公式 (8) 的条件 $\Delta c_{pq} \rightarrow 0^+$ 才能成立。结合图像求梯度时，使用了 3×3 大小窗口的 Sobel 算子^[15]。而 Sobel 算子属于一阶微分，所以中心像素为 3×3 大小的邻域，

实际直接相关的像素区域大小为 5×5 ，模型如图 1。

在同时满足计算速度与精度的情况下，结合 SSE 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^[15]。由此再根据得到的权重值，可以计算得到最大的窗口权重 w_{\max} 及对应的窗口形状如图 2。

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[15]所述，在统筹考虑算法的时间效率和误匹配率的前提下，作者测试了其它稀疏描述子的效果，最终选择 Sobel 作为特征描述子。但是，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，则破坏了公式(8)的 $\Delta c_{pq} \rightarrow 0^+$ 条件，无法使用提出的描述子窗口；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，对图像求取了两次导数，因此对灰度变化强烈区域特征描述能力强，丢失了一阶微分生成的纹理特征，不利于后续步骤中强鲁棒性匹配点的提取。而一阶微分算子中如 Roberts 算子卷积模板太小，对噪声敏感；而 Prewitt 算子中心权重比 Sobel 更小，微分后对图像过度平滑处理，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，微分后整幅图像上的平滑性更小。这使得在像素值相似的区域内，微分后的结果间的差异更大。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，使用平滑性更强的 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，为获得更小的平滑特性，增大窗口总体的特征描述，使用中心权重更大的 Scharr 滤波器可进一步降低误匹配。

源像素经过 Scharr 卷积内核后，增大了像素间的差异。特别是当中心点邻域内像素值差别过小时，Scharr 滤波器能提高中心权重，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，记这些点的组成的集合为 $\mathbf{S} = \{s_1, s_2, \dots, s_M\}$ ，其中每个支撑点都是一个三维列向量

$s_M = (u_m, v_m, d_m)^T$ ，它的定义为图像坐标 $(u_m, v_m) \in \mathbb{N}^2$ 和它的视差 $d_m \in \mathbb{N}$ 的组合。

假设观测值 $\{o_n^{(l)}, o_n^{(r)}\}$ 和支撑点 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。 $\mu(S, o_n^{(l)})$ 是连接支撑点和观测值的均值函数，能描述为线性分段函数，使用三角剖分根据支撑点计算出需要的内插视差值。对每个三角形构成的平面定义如下：

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

公式 (11) 中下标 i 表示像素点 (u_n, v_n) 所属的三角形， $o_n = (u_n, v_n, f_n)^T$ 是观测值。因此，文中提出的模型 μ 可以粗略的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

根据上述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，公式如下：

$$d_n^* = \arg \max p(d_n | o_n^{(l)}, o_1^{(r)}, o_2^{(r)}, \dots, o_N^{(r)}, S), \quad (12)$$

在右像平面沿着极线的观测值是结构化的，即给定与 $o_n^{(l)}$ 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，这使其具有非零概率。通过建模，沿着极线的所有观测值的分布可以获得这一特性：

$$p(o_1^{(r)}, o_2^{(r)}, \dots, o_N^{(r)} | o_n^{(l)}, d_n) \propto \sum_{i=1}^N p(o_i^{(r)} | o_n^{(l)}, d_n), \quad (13)$$

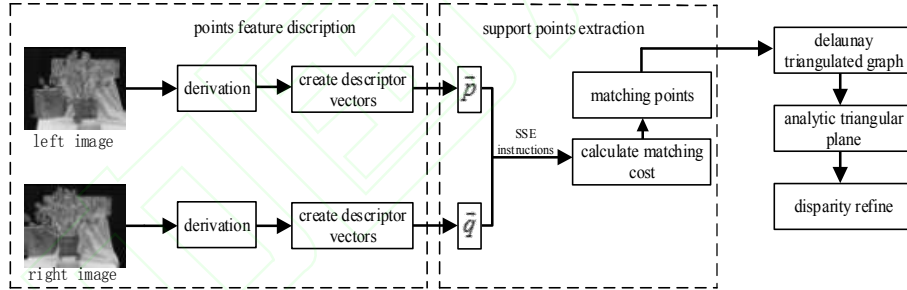


图 3 本文匹配主流程图

Fig. 3 Main flow chart matched in this paper

至此，整个算法的流程如图 3。图 3 中 points feature description 处理模块包含了本文提出的高斯窗口与 Scharr 滤波器，其中 Scharr 对应 derivation 处理流程，高斯窗口包含于 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了匹配代价的计算与匹配代价聚合，为此，本文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，为后续处理步骤提高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^[18]提供的第

三版标准测试数据集，数据集示例图见图 4，因为新数据集更具有挑战性且具有更大的半遮挡区域，评估网站将 nonocc (non-occluded, 非遮挡) 作为默认掩模 (本实验评估指标下的结果采用相同掩模获得)，而本节客观评价标准使用 Middlebury 平台的示例程序中的两项默认评估指标：误匹配率 (记为 bad) 与绝对值均差 (记为 avgErr, average absolute)。

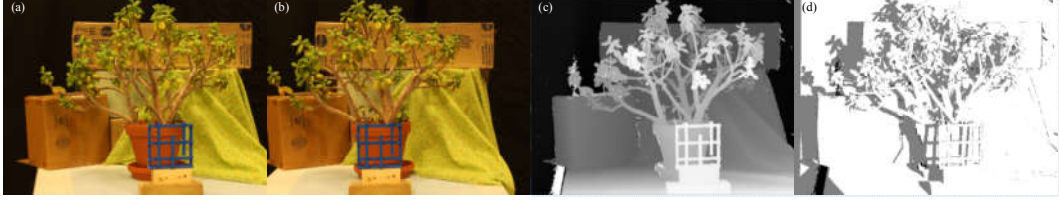


图 4 数据集示例图 (Jadeplant)。 (a)左图像; (b) 右图像; (c) 真实视差图; (d) 掩模

Fig. 4 Sample graphs of benchmark (Jadeplant). (a) Left image; (b) Right image; (c) Ground truth; (d) mask

评价指标 bad 的含义是，在掩模下设置误差限为 1，即当 ground truth 图和视差图结果相差 1 pixel 以上时，判定该点为误匹配点，公式如下：

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

公式 (14) 中 N 表示掩模下有效点的总和， n 表示超出误差限的匹配点总和。

评价指标 avgErr 的含义是，在掩模下计算非遮挡区域的 ground truth 值 gt 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，即其表示得到的视差值与 ground truth 值的平均总偏差的百分比，公式如下：

$$avgErr = \frac{\sum_i^N |d_i - gt_i|}{N - I_{\text{nf}}} \cdot 100\% , \quad (15)$$

其中 I_{nf} 表示视差值为 INFINITY 的无效点的总和。根据公式可见该项指标，在 bad 指标相近时更有意义，即脱离误匹配率，单一的对比 avgErr 指标并无明显的说服力，公式间的区别体现了该项指标与 bad 指标并不总是一一对应的关系。

$$SW = \sum_j^M (El_j - Pr_j) , \quad (16)$$

上式为误匹配及绝对值均差降低总量计算公式，其中 El 与 Pr 对应表示后文表中的 ELAS 列与 Proposed 列， M 表示对应列的行总数。

本节实验不仅测试了所提方法的有效性与其它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 Schar 滤波器结合的有效性。其次，将本文方法与典型传统算法的匹配性能对比；然后，对比了所提方法与新颖立体匹配算法的误匹配率，验证所提方法的有效性与适用性。最后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 C++，硬件环境 Intel® Core™ i3 CPU、8GB 内存，系统 CentOS7。

3.1 高斯权重窗口验证

实验结果表明，在只使用高斯权重窗口的情况下，对于 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 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表现性能相同，结果对比见表 1。

表 1 高斯权重窗口在两种指标下的性能

Table 1 Performance of Gaussian weight window under two metrics

数据名称	bad(%)		avgErr(%)		数据名称
	ELAS	Proposed	ELAS	Proposed	
Adirondack	12.07	11.89	0.88	0.877	Adirondack
ArtL	19.8	19.27	1.76	1.63	ArtL
Jadeplant	34.68	34.66	1.17	1.17	MotorcycleE
MotorcycleE	13.45	13.21	2.21	2.08	PianoL
PianoL	37.52	36.53	1.96	1.93	Pipes
Pipes	15.3	14.76	1.9	1.8	Playroom
Playroom	26.68	26.04	4.55	4.27	Playtable
Playtable	26.38	26.29	0.93	0.92	PlaytableP
PlaytableP	14.84	14.74	0.89	0.89	Recycle
Recycle	17.85	16.85	0.91	0.91	Teddy
Vintage	33.69	33.66	2.43	2.43	Vintage

经实验验证，高斯权重窗口降低了误匹配率，且在更多的测试组上取得了更优的表现。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 avgErr 指标计算的是掩模条件下的所有匹配点（包含误匹配点）与 ground truth 值差的绝对值的总和的百分比，它表示的是每一点与 ground truth 值的平均偏离程度，特别是误匹配点偏离大时会严重影响该百分比，因此该项性能并不总是与 bad 评价指标对应。实验对比表明高斯权重窗口更具有普适性，更能适应一般的场景切换。

为进一步验证高斯权重窗口的有效性，在图 5 的结果中，不仅通过实验验证了其它权重对应的窗口形状的不适用性，而且验证了不符合大小 5×5 窗口的不适用性（对应窗口形状见图 6）。通过分析多个窗口匹配结果，验证了高斯权重窗口的有效性，即高斯权重窗口能更有效的描述匹配点特征信息。其它窗口在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，最终致使误匹配的出现。

在分析图 5 的结果时，注意到这样一个事实：虽然不符合高斯权重的窗口误匹配率高，在测试集上大部分实验组的性能不佳，但是在个别图像上一些形状的窗口性能取得了更优的表现，如 D 型窗口，对 Jadeplant 图像的误匹配率低于其它窗口，但是在其它图像上的误匹配率更高。分析原因，所提窗口能应对更多的场景类型，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，特殊类型窗口能更好的描述特征信息。

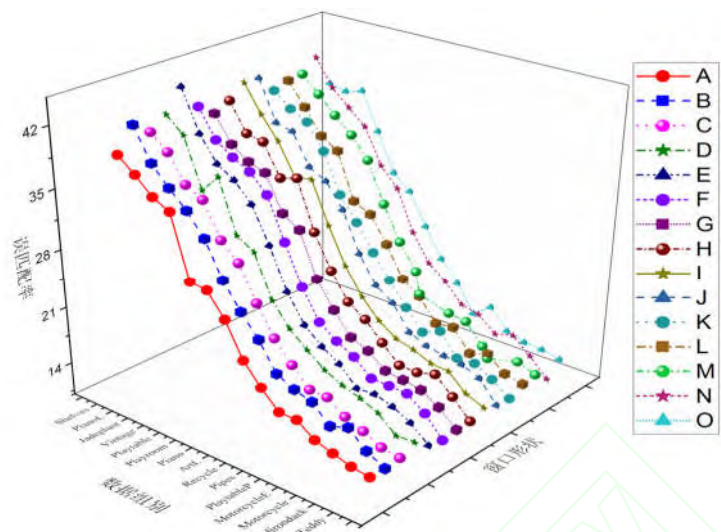


图 5 不同窗口形状误匹配率

Fig. 5 Mismatching rate of different window shapes

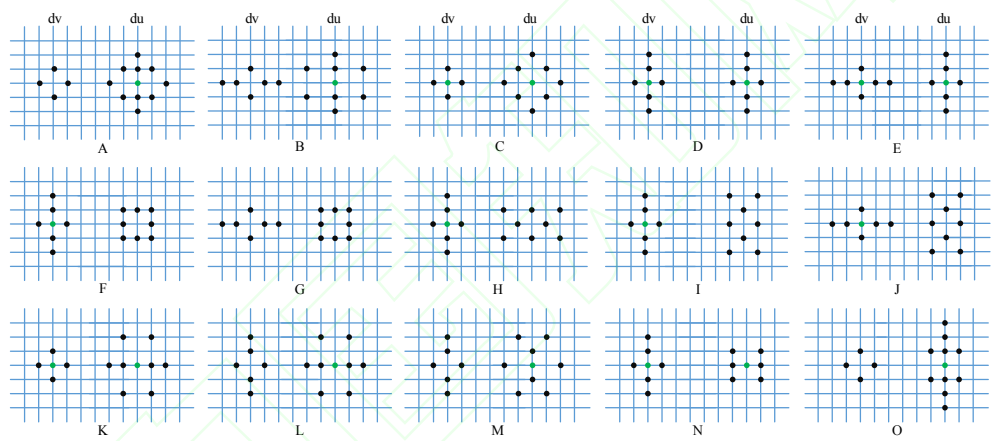


图 6 不同类型的匹配窗口

Fig. 6 Different types of matching windows

图 6 中，A 窗口是符合本文原理的窗口，中心点表示取点两次(考虑该点权重大次最大权重值的近二倍)。实验中修改了 dv 和 du 的点数和形状（总点数保持为 16），证明了其它窗口的无效性。类似于图中标号为 O 的窗口类型，其与本文使用的窗口（A）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 形状的窗口，本文还测试了其它形状，其总体性能并不如本文提出的窗口类型，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，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，能提高邻域内匹配点的特征描述，为匹配代价的计算提供了更有力的支撑。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，对比结果见表 2。

Scharr 滤波器在 bad 评价指标下取得了优势性能，而在 avgErr 评价指标下表现出了相对劣势，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了中心权重，适当缩小了像素间的平滑性，但是改进并未改变 Sobel 算子微分的性质，而是提升了 Sobel 算子在数字图像上的性能。综合两个评价指标，Scharr 滤波器总体上取得了优于原算法的性能，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

Table 2 Performance of Scharr filter under two metrics

数据名称	bad(%)		avgErr(%)		数据名称
	ELAS	Proposed	ELAS	Proposed	
ArtL	19.8	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	37.5	1.96	1.97	Pipes
Pipes	15.3	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

在上述实验基础上，本小节目的之一是验证高斯权重窗口与 Scharr 滤波器结合后的性能(在后文中，未经特殊说明，“本文方法”均指两者结合的方法)，通过增强的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。实验结果表明，本文方法在 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，对比结果见表 3。

表 3 本文方法在两种指标下的匹配性能

Table 3 Performance of proposed algorithm under two metrics

数据名称	bad(%)		avgErr(%)		数据名称
	ELAS	Proposed	ELAS	Proposed	
Adirondack	12.07	11.81	0.88	0.83	Adirondack
ArtL	19.8	19.55	1.76	1.74	ArtL
Jadeplant	34.68	27.85	8.95	5.99	Jadeplant
MotorcycleE	13.45	13.3	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	15.3	15.3	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

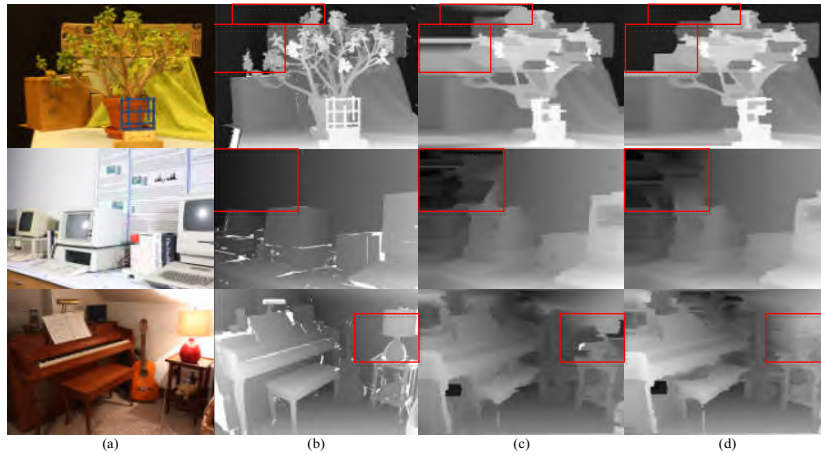


图 7 本文方法与原方法效果提升示例。(a) 左图像；(b) 真实视差图；(c) ELAS；(d) 本文算法

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

经实验验证，本文方法不仅在 bad 指标下性能优于单独使用高斯权窗口和 Scharr 滤波器（误匹配率总降低量分别为 12.85%、4.36%和 7.88%），而且在 avgErr 指标下的性能同样优于单独使用高斯权窗口和 Scharr 滤波器（绝对值均差总降低量分别为 3.53%、0.683%和 2.71%）。图 7 给出了本节实验的效果提升的示例图（三行分别为 Jadeplant、Vintage 和 PianoL），方框标注了本文方法在视差图上的提升效果。如在 Vintage 与 Jadeplant 场景中，本文视差图的条纹明显减少，纹理性更好；在 PianoL 的场景中，本文方法产生的零值区域更少。达到了降低误匹配率，改善视差图中的条纹及空洞现象。

本文方法在单对图像上的运行时间最多降低 0.03s，总运行时间平均降低 0.0033 秒。可见本文方法几乎不影响算法的运行时间，即本文方法并未试图牺牲匹配效率提高匹配率。

在相同分辨率下，图 8（评价指标 bad）和图 9（评价指标 avgErr）将本小节实验结果与新颖算法和传统算法做了对比，其中 GwcNet^[19]和 GA-Net^[20]均为基于 CNN 的新颖立体匹配算法，而传统匹配算法则使用局部立体匹配算法 Census 与半全局立体匹配算法 SGM。

综合分析测试的结果，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；然而本文方法优于这两个新颖方法（除 Playtable 图像对外），分析认为在测试时使用了预先训练好的模型，而测试数据非 KITTI 数据集，如前文所述，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匹配率明显下降，然而本文方法不依赖于训练模型，取得了更优的性能表现；其次本文方法在 PlaytableP 的图像对上的匹配性能要高于在 Playtable 上的，原因在于前者为完美校正（perfect rectification）的图像

对，而本文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 均使用了互信息，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 值略低于两个新颖算法，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本文方法，但在整体上，本文方法的该项指标取得了绝对优势（10/15 的该项性能更优），另外这四组图像对的 bad 指标明显低于本文方法，在 avgErr 值近似时误匹配率低更有说服力。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本文方法，本文方法并不依赖于某一特定场景的特征信息，因而能面对各类视觉场景，更符合机器视觉发展的一般要求。

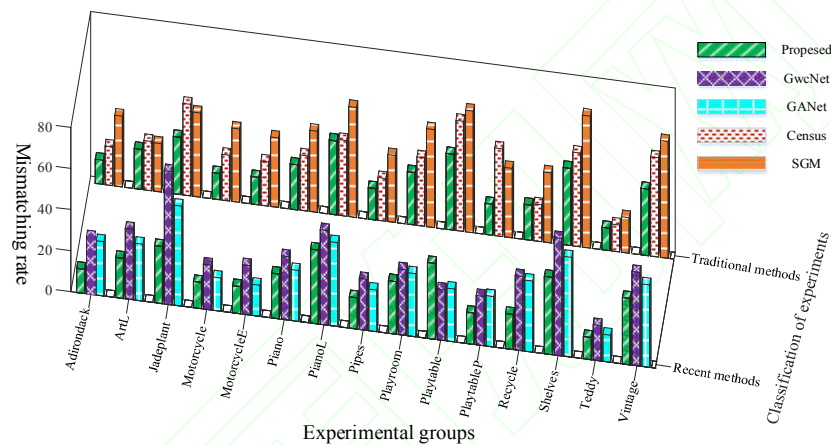


图 8 bad 指标对比图

Fig. 8 bad metric comparison chart

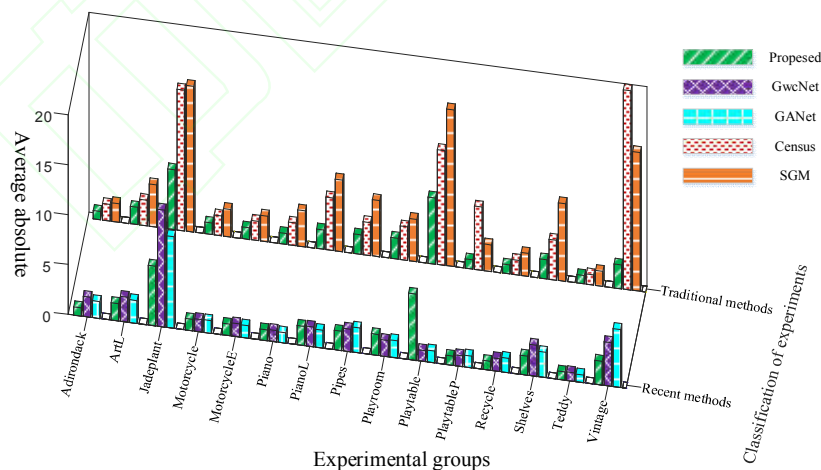


图 9 avgErr 指标对比图

Fig. 9 avgErr metric comparison chart

本文方法同样优于测试的两个传统算法的匹配性能，这是因为本实验使用的数据集更具有挑战性，即病态条件要更严重，而 SGM 聚合特定方向上的全局信息，导致病态区域的信息不同程度地扩散到整幅图像上，影响了 SGM 的匹配性能；而 Census 变换和本文方法在计

算匹配代价时都使用大小一定的匹配窗口聚合信息，抑制了病态信息的传播。但 Census 变换只保留窗口内像素之间的关系，造成原图像灰度信息的丢失，这在灰度值近似区域容易造成误匹配，而本文的方法避免了这一现象，获得了更高的匹配率。

在主观上，视差图对比示例见图 10（三行分别为 Adirondack、Jadeplant 和 Shelves 图像），方框内标识了本文方法明显提升的效果。其中 GA-Net 视差图与 ground truth 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细致纹理。对于 Census 视差图，其遮挡区域出现大面积条纹，视差图纹理不清晰且有不连续区域；而 SGM 的视差图，在前景物体上出现了零值区域，物体形状出现变形（如 Jadeplant 视差图），且 Shelves 的视差图纹理不清晰。

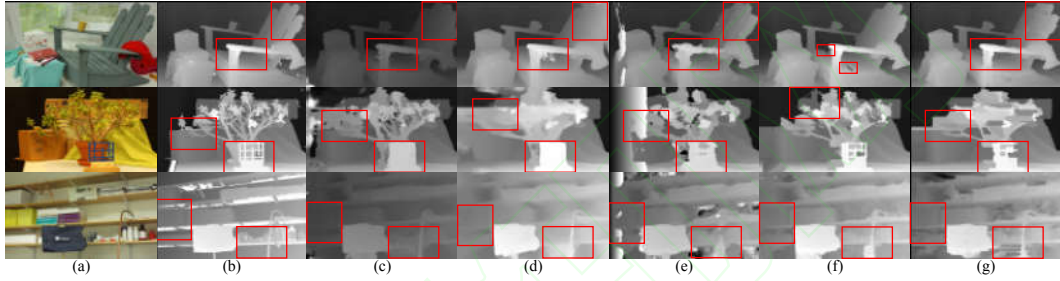


图 10 提升效果对比示例图。(a) 左图像；(b) 真实视差图；(c) GA-Net；(d) GwcNet；(e) Census；(f) SGM；(g) 本文算法

Fig. 10 Sample graphs of effectiveness improvement. (a) Left images; (b) Ground truth; (c) GA-Net; (d) GwcNet; (f) SGM; (g) Proposed

3.4 真实场景与水下图像测试

为进一步验证本文方法的鲁棒性，本小节测试了自采集场景图，由于缺少 ground truth 图，实验数据无法客观对比，只能主观上对比算法间的优劣，结果见图 11（三行分别为 Air_1、Air_2 和 Underwater 图像）。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， d 为视差值）和 ground truth 图特点，获得以下几点主观评价标准：

- 1) 纹理清晰，视差连续；
- 2) 室内场景无视差值为零的区域；
- 3) 前景物体的视差值要比背景的大。

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）。而本文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 也表现出了这些特点，但是如

Air_1 对应的各算法视差图，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，而本文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的数值更大。对于 Air_2 对应的视差图，GwcNet 与本文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 与本文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；这一点本文方法虽然产生了小部分零值区域，而在其遮挡区域的视差值比柱子的视差值低，更符合实际情况的情况。

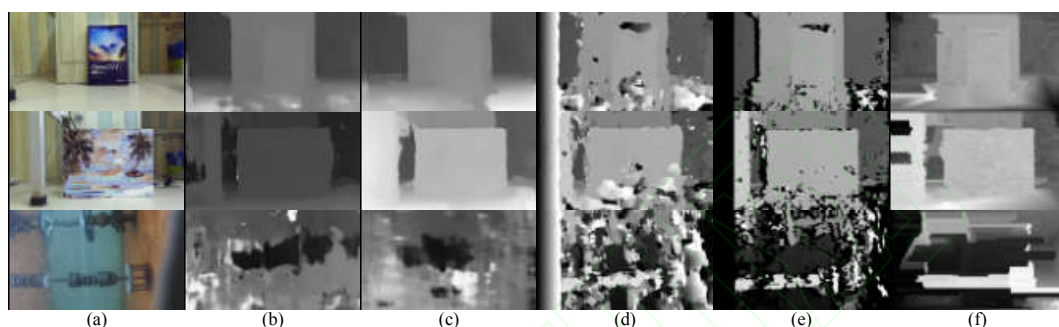


图 11 自采集图像。(a) 左原图；(b) GA-Net；(c) GwcNet；(d) Census；(e) SGM；(f) 本文方法
Fig. 11 Self-collecting images. (a) Left images; (b) GA-Net; (c) GwcNet; (d) Census; (e) SGM; (f) Proposed

水下图像（见图 11 的 Underwater）由于吸收、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，都表现出了很低的匹配性能。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，但是前景物体无大致轮廓，前景与背景间也无明显的层次感。在本文方法生成的视差图中，虽然前景物体也无清晰轮廓，但是具有两个前景物体的大致轮廓，与背景间有明显的层次感，而且视差图连续性更少，并且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，本文方法的鲁棒性更高，更能面对复杂的场景，并且不受场景切换的影响。

5 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问题的目的。同时利用待匹配点与临近像素点的几何信息，和像素点微分信息，增强了临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了明显改善，与原算法相比误匹配率降低总和为 12.85%（10/15 组的优势性能），且 avgErr 指标降低总和为 3.53%（11/15 组的优势性能），误匹配率等有效降低。此外，本文算法作为一种增强待匹配点特征信息的算法，可以容易的与其它立体匹配算法结合，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]//IEEE Conference on Computer Vision and Pattern Recognition. 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]//IEEE International Conference on Computer Vision. IEEE, 2017: 66-75.
- [3] Chang J R, Chen Y S. Pyramid stereo matching network[C]//IEEE Conference on Computer Vision and Pattern Recognition. 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(02): 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(02): 021003.
- [6] Zhang Y, Khamis S, Rhemann C, *et al.* Activestereonet: end-to-end self-supervised learning for active stereo systems[C]//European Conference on Computer Vision. Springer, 2018: 784-801.
- [7] Žbontar J, LeCun Y. Computing the stereo matching cost with a convolutional neural network[C]//IEEE conference on computer vision and pattern recognition. IEEE, 2015: 1592-1599.
- [8] Xiao J S, Tian H, Zou W T. Stereo matching based on convolutional neural network[J]. Acta Optica Sinica, 2018,38(08): 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(08): 0815017.
- [9] Spyropoulos A, Komodakis N, Mordohai P. Learning to detect ground control points for improving the accuracy of stereo matching[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014: 1621-1628.
- [10] Tani T, Matsushita Y, Naemura T. Graph cut based continuous stereo matching using locally shared labels[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014: 1613-1620.
- [11] Sinha S N, Scharstein D, Szeliski R. Efficient high-resolution stereo matching using local plane sweeps[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014: 1582-1589.
- [12] Zhang K, Fang Y, Min D, *et al.* Cross-scale cost aggregation for stereo matching[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014: 1590-1597.
- [13] Pang J, Sun W, Yang C, *et al.* Zoom and learn: Generalizing deep stereo matching to novel domains[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 2070-2079.
- [14] Tonioni A, Tosi F, Poggi M, *et al.* Real-time self-adaptive deep stereo[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 195-204.

- [15] Geiger A, Roser M, Urtasun R. Efficient large-scale stereo matching[C]//Asian conference on computer vision. Springer, 2010: 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(02): 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类方法[J]. 机器人, 2018, 40(02): 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. Middlebury stereo vision page(2014-10-15)[EB/OL]. [2019-08-26]. <http://vision.middlebury.edu/stereo/>.
- [19] Guo X, Yang K, Yang W, *et al.* Group-wise correlation stereo network[C]//IEEE Conference on Computer Vision and Pattern Recognition. 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]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 185-194.

网络首发:

标题: 改进窗口特征及微分算子的立体匹配算法

作者: 李新春,殷新勇,林森

收稿日期: 2019-08-29

录用日期: 2019-09-24

DOI: 10.3788/lop57.081504

引用格式:

李新春,殷新勇,林森. 改进窗口特征及微分算子的立体匹配算法[J]. 激光与光电子学进展, 2020, 57(08):081504.

网络首发文章内容与正式出版的有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

基于机器视觉的包装袋缺陷检测算法研究与应用

李丹 白国君 金媛媛 童艳

沈阳城市建设学院信息与控制工程系, 辽宁 沈阳 110167

激光与光电子学进展, 2019, 56(9):091501

基因表达式编程优化的色调保持低照度图像增强

贾新宇 李婷婷 江朝晖 刘海秋 饶元

安徽农业大学信息与计算机学院, 安徽 合肥 230036

激光与光电子学进展, 2019, 56(9):091502

基于深度学习的图像显著区域检测

纪超 黄新波 曹雯 朱永灿 张烨

西安工程大学电子信息学院, 陕西 西安 710048

激光与光电子学进展, 2019, 56(9):091007

基于二次引导滤波的局部立体匹配算法

王凯 李志伟 朱成德 王鹿 黄润才 郭亨长

上海工程技术大学电子电气工程学院, 上海 201600

激光与光电子学进展, 2019, 56(8):081004

面向皮革缺陷检测的离轴LED曲面阵列照明设计

王德宇 王雪琨 于文文 曹建军 钱维莹 高淑梅

江南大学理学院江苏省轻工光电工程技术研究中心, 江苏 无锡 214122

激光与光电子学进展, 2019, 56(8):082202