

双目视觉的匹配算法综述

陈 炎, 杨丽丽, 王振鹏

(中国农业大学信息与电气工程学院, 北京 100083)

摘 要: 双目视觉是获取对现实世界立体感知的重要方法, 在自动驾驶等领域得到了普遍的应用。立体匹配是实现双目感知的前提, 该算法对左右摄像机拍摄的照片进行像素级的匹配, 生成稠密视差图, 从而获取了三维坐标信息。概述了立体匹配算法近 20 年来的发展过程, 围绕基于人工特征和深度学习两个方向进行了综述, 对算法实现过程中的代价计算、代价聚合、视差计算和视差求精进行分析讨论, 评估了算法的准确性和时间复杂度。最后总结了立体匹配算法面临的挑战和对未来发展的展望。

关 键 词: 立体匹配; 双目视觉; 立体感知; 深度学习; 计算机视觉

中图分类号: TP 391.4

DOI: 10.11996/JGj.2095-302X.2020050702

文献标识码: A

文章编号: 2095-302X(2020)05-0702-07

Literature survey on stereo vision matching algorithms

CHEN Yan, YANG Li-li, WANG Zhen-peng

(College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Stereo vision is an important means to acquire stereoscopic perception of the real world, which is widely used in autonomous driving and other fields. Stereo matching algorithm serves as the basis for the binocular system to realize perception. This algorithm matches the images captured by the left and right cameras and calculates the depth map, thus providing coordinates for the 3D modeling. There are four stages of stereo matching algorithm, including cost computation, cost aggregation, disparity selection, and disparity refinement. This literature survey summarized the development of stereo matching algorithm from such two research fields as artificial feature and deep learning in the past 20 years, evaluated the accuracy and time complexity, and presented the challenges and prospects of the development of stereo matching algorithms.

Keywords: stereo matching; stereo vision; stereo perception; deep learning; computer vision

立体视觉是机器认识世界的重要手段。利用 2 幅图像形成的视差可以计算得到像素点的深度, 从而获得三维的立体感知^[1-2]。其工作流程主

要包含 4 步: 双目标定、图像校正、立体匹配和三维重建, 而立体匹配是关键步骤^[3]。

立体匹配算法可以实现 2 幅图像像素级的匹

收稿日期: 2020-04-11; 定稿日期: 2020-05-28

Received: 11 April, 2020; Finalized: 28 May, 2020

基金项目: 国家重点研发计划项目(2016YFB0501805)

Foundation items: National Key Research and Development Programs of China (2016YFB0501805)

第一作者: 陈 炎(1994-), 男, 河北衡水人, 硕士研究生。主要研究方向为机器视觉。E-mail: 13122358665@163.com

First author: CHEN Yan (1994-), male, master student. His main research interests cover machine vision. E-mail: 13122358665@163.com

通信作者: 杨丽丽(1974-), 女, 内蒙古赤峰人, 副教授, 博士。主要研究方向计算机网络与智能信息处理。E-mail: llyang@cau.edu.cn

Corresponding author: YANG Li-li (1974-), female, associate professor, Ph.D. Her main research interests cover computer networks and intelligent information processing. E-mail: llyang@cau.edu.cn

配,在该过程中,由于存在光照不均匀、遮挡、模糊以及噪声等的影响,给提升匹配精度带来了巨大的挑战^[4]。在双目视觉的应用中,如自动驾驶、机器人等领域^[5-6],需要具备实时、准确的深度估计,因此对于系统的处理速度和精度提出了很高的要求。在以往的综述性文章中^[3,7-9],研究者们对基于人工特征算法的原理及性能进行了总结,近年出现了许多基于深度学习的优秀算法,其在实时性和准确性上都有很大的提升。为了给相关研究人员提供参考,本文讨论了各种立体匹配算法的发展,对基于深度学习和人工特征的算法特点加以分析,并给出了立体匹配算法的发展方向。

1 基于人工特征的匹配算法

立体匹配算法实现流程主要分为4步:代价计算、代价聚合、视差计算和视差求精^[7],流程如图1所示。代价计算是对匹配区域像素点和参考区域进行相似度计算。由双目相机的成像原理可知,人们只需要计算处于同一水平线上的左右图像的像素相似值^[10],这一要求基于对极约束^[11]来实现。极约束在立体匹配中起重要作用,其将对应关系的搜索限制为一行而不是整个图像空间,从而减少了所需的时间和搜索范围。代价聚合阶段通过设置能量函数,将中心像素点的相似度值用某个范围内的像素点代价计算结果替代。常用的相似性评价方法包括绝对差之和(sum of absolute difference, SAD)、平方差之和(sum of squared differences, SSD)、归一化互相关(normalized cross correlation, NCC)、秩变换(rank transform, RT)和普查变换(census transform, CT)^[9]。视差计算阶段,可选取参考区域内相似度最高的像素点来计算。视差求精的目的是减少错误匹配的像素点,优化视差图。优化步骤包括正则化和遮挡填充或插值^[12]。

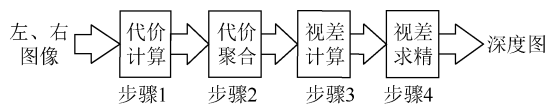


图1 立体匹配算法流程

Fig. 1 The framework of stereo matching algorithm

按照代价函数约束范围的差别,立体匹配算法可分为全局匹配法和局部匹配法^[8]。

1.1 全局立体匹配算法

全局匹配法的能量函数整合了图像中的所有像素,以尽可能多地获取全局信息。函数的表达

式为

$$E_d = E_{\text{data}}(D) + \lambda \cdot E_{\text{smooth}}(D) \quad (1)$$

其中, $E_{\text{data}}(D)$ 为数据项,表示全部像素的匹配代价; $E_{\text{smooth}}(D)$ 为平滑项,表示相邻像素对视差值的一致性; λ 为权值参数,取正值; E_d 为全局能量函数。根据 E_d 优化方法的不同,全局匹配法又可分为动态规划法(dynamic programming, DP)、置信度传播法(belief propagation, BP)和图割法(graph cut, GC)^[10]。

SUNG等^[13]在求解 E_d 的过程中,应用极线间的相关性,构建了多路径自适应的动态规划求解方法。LI等^[14]为了缓解动态规划法引起的横条纹效应,从左侧和右侧图像分别提取改进的SIFT算法描述作为特征点完成匹配,有效地减少了这一现象。LEUNG等^[15]使用四叉树对动态规划进行快速迭代,提高了算法的运行速度。BLEYER和GELAUTZ^[16]使用图割法,假设深度均匀变化,且深度变化边界与图像特征边界重合,在初始视差段聚集形成一组视差层,通过全局成本函数的最小化选取最佳视差层,在大规模无纹理区域取得了良好的匹配效果。LEMPITSKY等^[17]设计了多重标记的马尔科夫随机域图形分割,提高了运行速度。WANG等^[18]将图形分割后的点分为可靠和不可靠点,在遮挡和无纹理区域的不可靠区域,使用可靠区域的迭代信息进行优化,结果比经典图割法效率提高了90%。

1.2 局部立体匹配算法

局部匹配法将参考图像分为若干图像块,再求取匹配图像内预期相似度最高的图像块,生成深度图^[19]。局部匹配法与全局匹配法相比,能量函数只有数据项,而没有平滑项,因此只能求取局部最优解^[20]。算法原理如图2所示,随机选取左图像中 $m \times n$ 的待匹配图像块,搜索右图像视差 $(-d, d)$ 的范围,选取相似度最大的区域作为匹配单元。

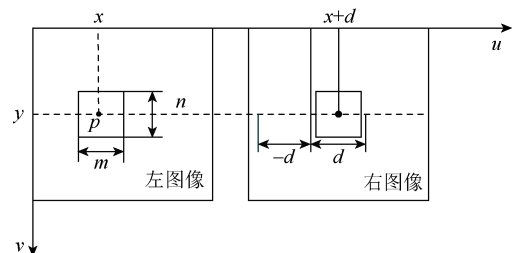


图2 局部立体匹配算法

Fig. 2 Local stereo matching algorithm

局部匹配法带来了效率上的提升,但由于缺乏

全局参数,生成的视差图不甚理想。为了提升匹配效果,在代价计算阶段可以使用更优的方法。**HIRSCHMULLER**^[21]提出了半全局匹配方法,主要包含2点:一是基于互信息匹配的分层计算,二是使用一维最优求解全局能量函数,半全局匹配在保证算法运行效率的同时大大提升了匹配效果。**WANG**和**ZHENG**^[22]在构建局部能量函数的过程中引入区域间的合作竞争机制,获得了良好的视差效果。**HOSNI**等^[23]提出在局部匹配时采用分割算法选取待匹配区域,可以提升视差图的质量,在分割选取待匹配区域时,以距离中心像素的远近分配不同的权重,改善了分割效果。**YOON**等^[24]针对各分割区域,提出了一个非线性的基于扩散的权重聚合方法,有效改善了遮挡区域匹配效果。**ZHANG**等^[25]提出自适应匹配的概念,并开发了具有任意自适应形状的支撑区域,实现了很高的匹配精度。

局部匹配算法在提升效果的同时,不可避免的会带来效率上的降低。为此,**DI**等^[26]在匹配中引入特征点选择,为了进一步提高运算效率,匹配图像中的特征点进行WTA匹配,非特征点只进行简单验证,提高了边缘的匹配效果。**MATSUO**等^[27]在匹配成本计算阶段使用基于AD算法和Sobel算子的局部方法,通过多次迭代和JBF固定窗口生成了精确的视差图。

2 基于深度学习的匹配算法

手工设计的人工特征,缺乏对上下文信息的获取,经验参数的选择对匹配效果影响很大,不适合在复杂环境下应用。深度学习通过卷积、池化、全连接等操作,对图像进行非线性变换,可以提取图像的多层特征用于代价计算,对提取的图像特征进行上采样过程中设置代价聚合和图像增强方法,从而实现端到端的图像匹配。深度学习的方法更多地利用了图像的全局信息,通过预训练获得模型参数,提高了算法的鲁棒性。用于立体匹配的图像网络主要可分为金字塔网络、孪生网络(siamese network)和生成对抗网络。

2.1 图像金字塔网络

为了保证输入图片尺寸的一致性,常常需要对图片进行裁剪、缩放等操作。为了解决这个问题,**HE**等^[28]在卷积层和全连接层之间设置空间金字塔池化层,将不同尺寸的图片特征转化为固定长度的表示,避免了卷积的重复计算,在图像分割、图像

匹配等领域获得了极大的提升效果。**ŽBONTAR**和**LECUN**^[29]首次提出使用卷积神经网络提取图像特征用于代价计算,设置跨成本交叉的代价聚合,运用左右一致性检查消除错误的匹配区域,其标志着深度学习开始成为立体匹配的重要手段。**CHANG**和**CHEN**^[30]将金字塔池化模块引入到特征提取中,使用多尺度分析和3D-CNN结构,有效地解决了梯度消失和梯度爆炸的问题,在弱纹理、遮挡、光照不均匀等条件下获得了良好的效果。**DUGGAL**等^[31]提出一种全新可微的Patch Match算法。首先获得稀疏的视差图,以减少代价计算的复杂度,再推算出剩余像素点的匹配范围,并在KITTI和SceneFlow数据集上实现了每对图片62 ms的匹配速度。**TONIONI**等^[32]构建了模块化、轻量化的神经网络,网络子部分可独立训练,使用无监督算法模型,使网络可以不断更新,在不降低质量的前提下,达到了每秒40帧的处理速度。**GUO**等^[33]构建了分组代价计算,**ZHANG**等^[34]设计了半全局聚合层和局部引导聚合层,都可以代替3D卷积层,提高计算效率。

2.2 孪生网络

孪生网络的概念最早由**BROMLEY**等^[35]提出,其基本结构如图3所示。通过设置2个权重共享的卷积神经网络,分别输入左、右图像,通过映射函数将卷积得到的特征转化为特征向量,衡量2个特征向量间的L1距离就可以拟合2张图片的相似度。文献[36]对原有的孪生网络进行了改进,使用RELU函数和小卷积核加深卷积层,提高了匹配精度。文献[37]使用孪生网络从左右图像提取特征,首先在低分辨率的代价卷积中计算视差图,再使用分层细化网络引入高频细节,利用颜色输入作为指导,可以生成高质量的边界。**LIU**等^[38]提出通过连接2个子网络获取多尺度特征。第1个子网络由孪生网络和3D卷积网络构成,可以生成低精度的视差图;

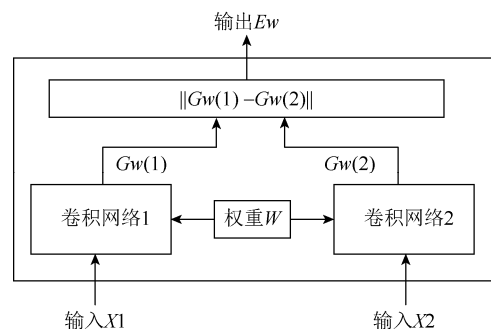


图3 孪生网络结构

Fig. 3 Siamese network structure

第2个子网络是全卷积网络,将初始视差图恢复成原始分辨率,2个网络用金字塔池化进行连接。文献[39]在低分辨率的视差图上对深度不连续区域进行改进,在视差求精阶段恢复成原始分辨率,取得了良好的效果。

2.3 生成对抗网络

在博弈论的基础上,文献[40]构建了生成式对抗网络(generative adversarial networks, GAN)。网络由生成模型和判别模型构成。生成模型学习样本特征,生成图像与原图像相似,而判别模型则用来分辨“生成”图片与真实图片。这个过程不断迭代运行,最终判别结果达到纳什均衡,即真假概念均为0.5。PILZER等^[41]提出了基于双目视觉的GAN框架,其由2个生成子网络和1个判别网络构成。2个生成网络在对抗学习中分别被用来训练重建视差图,通过相互制约和监督,生成2个不同视角的视差图,融合后输出最终数据。实验表明,这种无监督模型在光照不均匀的条件下可以得到良好的效果。文献[42]使用生成模型对遮挡区域进行处理,恢复得到了良好的视差效果。文献[43]提出了深卷积生成对抗模型,可以通过相邻帧获得多幅深度图,进一步改进了遮挡区域的深度图效果。文献[44]通过左右相机拍摄的2幅图像,生成一幅全新的图像,用于改进视差图中匹配效果较差的部分,实验表明,该方法对于光照强度较差的区域视差图提升效果明显。

3 算法测试与评估

3.1 算法测试框架

目前,立体匹配算法主要有Middlebury和KITTI 2个测评数据集,主要对算法的误匹配率和运行时间进行准确的测评,并提供算法的排名。这2个数据集都构建了标准的视差图,可用于计算误匹配率。KITTI提供了统一的云平台运行算法,保

证了算法运行时间测评的硬件一致性。

Middlebury^[7,45-48]由卡内基梅隆大学双目实验室创建,评价系统可以求取整张图片、遮蔽和深度不连续区域的误匹配率,并对提交的算法进行横向对比排名。

2011年,德国的卡尔斯鲁厄理工学院和芝加哥丰田技术研究所联合创立了Kitti立体匹配测试平台,并在2015年进行了更新^[49-50]。数据集包含车载相机采集的市区、乡村和高速公路图像,旨在为自动驾驶提供技术准备和评估。

3.2 算法测试指标

目前,立体匹配算法的主要评价标准为视差图精度和时间复杂度。视差图精度的评价指标有误匹配率,平均绝对误差和均方误差。误匹配率的计算式为

$$B = \frac{1}{N} \sum (x, y) (|d_c(x, y) - d_{GT}(x, y)| > \delta_d) \quad (2)$$

其中, $d_c(x, y)$, $d_{GT}(x, y)$ 分别为生成视差图和真实视差图的像素值; δ_d 为评价阈值需设置,当差值大于 δ_d 时,此像素记为误匹配像素; N 为视差图的像素总数。平均绝对误差式为

$$AVE = \frac{1}{N} \sum (x, y) |d_c(x, y) - d_{GT}(x, y)| \quad (3)$$

均方误差的计算式为

$$RMS = \left(\frac{1}{N} \sum (x, y) |d_c(x, y) - d_{GT}(x, y)|^2 \right)^{\frac{1}{2}} \quad (4)$$

其中, N 为所有像素的总数和非遮挡区域的像素数量。

部分算法的评价结果见表1。

3.3 2种不同类型的算法评估

基于人工特征的立体匹配算法,将数学原理的发展广泛应用于实际需求,推动了双目视觉系统在各行各业的应用。其拥有完整的数学模型和方便可调的参数,模型的可调整性和堆叠性较好,在鲁棒

表1 部分立体匹配算法评价结果

Table 1 Some stereo matching algorithm evaluating results

类型	文献	图像宽高比	计算平台	误匹配率(%)	平均运行时间(ms)
基于人工特征	[16]	450×350	CPU AMD Phenom II 2.8 GHz	7.85	526
	[17]	450×350	CPU Intel Core 2.20GHz and GPU GTX480	5.43	2 530
	[23]	2864×2008	CPU Intel i5 2.6 GH	9.76	6 350
	[26]	450×350	Stratix II GX FPG	8.97	920
基于深度学习	[30]	1242×375	GPU GTX 1080Ti	2.21	420
	[32]	1242×375	GPU Titan-Xp	2.32	25
	[38]	1242×375	GPU Titan X Pascal	3.91	97

性要求不高的应用场景中,通过多次实验,可以得到良好的视差效果^[18,51-52]。

基于深度学习的立体匹配算法,可以提取更多的图像特征用于代价计算。相比于人工特征,深度学习可以获取更多的上下文信息,提高算法的鲁棒性,使用 GPU 加速技术,得到更快的处理速度^[30]。在自动驾驶、机器人等领域,深度学习通过大量数据对网络进行训练,可以获得高精度的视差图,满足了精度和实时性的要求^[32]。

人工特征中的局部匹配算法误匹配率较高,一般在 10%~30%之间^[7],运行时间上最快的 BM 算法处理一对 1432×1004 大小的 RGB 图片只需 0.05 s,误匹配率为 32.45%^[23],精度和速度平衡性较好的

SGBM 算法处理一对相同大小图片用时大约为 0.53 s,误匹配率为 18.36%^[21]。全局匹配算法的精度比局部匹配法要高,误匹配率一般在 10%左右^[7],但运行速度较低,处理图像的时间较长。KITTI 平台中的精度排名最高的全局匹配算法误匹配率为 5.12%,运行时间达到 65 s。基于深度学习的匹配算法通过大量数据训练可以达到很高的精度,KITTI 平台中精度排名最高的深度学习算法误匹配率只有 1.41%,在仅使用单核 CPU 运算的条件下运行速度为 0.52 s,表现出深度学习强大的应用前景。

对于基于人工特征和深度学习的匹配算法的特点总结见表 2。

表 2 算法特点对比
Table 2 Comparison of algorithm features

算法分类		优点	缺点	应用场景
基于人工特征的 立体匹配算法	局部匹配法	速度较快,模型堆叠性好	精度较低	机械臂抓取,工件测量,农田作物定位等
	全局匹配法	精度较高	时间复杂度高	三维建模,基于视觉地图的构建等
基于深度学习的立体匹配算法		实时性和对环境的适应性好	需数据预训练,硬件要求高	自动驾驶,机器人感知等

4 结 论

立体匹配是实现双目视觉的核心点,匹配的效果直接影响后续的目标识别、三维建模等工作。

基于人工提取特征的立体匹配算法由于数学解释性高,堆叠性好等优点,目前广泛应用于固定场景的识别。通过多种方法进行融合,可以取得良好的效果。

基于深度学习的立体匹配算法已成为自动驾驶、机器人领域的主流,目前要解决的难题主要分为以下 3 个方面:

(1) 提高算法的实时性和准确性。对于自动驾驶领域尤为重要,但目前算法在这 2 方面很难做到兼顾。

(2) 增强算法的鲁棒性。深度学习算法受限于数据集的训练,这使得算法对于农田、山地等陌生环境的匹配准确率有所降低。

(3) 提高算法的适应性。现实世界中雨雾天气下拍摄的图像可能存在反光、模糊、噪声等各种问题,这对算法提出了较高的要求。

人工特征的应用有助于深度学习网络的完善,而近年来,强化学习、图神经网络等领域的不断发展,也都为立体匹配算法的改进提供了参考。

参 考 文 献

References

- [1] LEMAIRE T, BERGER C, JUNG I, et al. Vision-based SLAM: stereo and monocular approaches[J]. International Journal of Computer Vision, 2007, 74(3): 343-364.
- [2] 亢宇欣, 谌贵辉, 邓宇, 等. 多测度融合的立体匹配算法研究[J]. 图学学报, 2019, 40(4): 711-717.
KANG Y X, CHEN G H, DENG Y, et al. Research on stereo matching algorithms based on multi-measure fusion[J]. Journal of Graphics, 2019, 40(4): 711-717 (in Chinese).
- [3] TIPPETTS B, LEE D, LILLYWHITE K, et al. Review of stereo vision algorithms and their suitability for resource-limited systems[J]. Journal of Real-Time Image Processing, 2016, 11(1): 5-25.
- [4] HONG P N, AHN C W. Stereo matching methods for imperfectly rectified stereo images[J]. Symmetry, 2019, 11(4): 570-585.
- [5] NALPANTIDIS L, GASTERATOS A. Stereovision-based fuzzy obstacle avoidance method[J]. International Journal of Humanoid Robotics, 2011, 8(1): 169-183.
- [6] 罗昌平, 夏海波, 赵琼. 基于运动视差的三维重建技术在自导引小车中的应用[J]. 图学学报, 2017, 38(4): 615-622.
LUO C P, XIA H B, ZHAO Q. The application of three-dimensional reconstruction technique based on motion parallax in AGV[J]. Journal of Graphics, 2017,

- 38(4): 615-622 (in Chinese).
- [7] SCHARSTEIN D, SZELISKI R, ZABIH R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. *International Journal of Computer Vision*, 2002, 47(1-3): 7-42.
 - [8] HAMZAH R A, IBRAHIM H. Literature survey on stereo vision disparity map algorithms[J]. *Journal of Sensors*, 2016, 45(3): 1-23.
 - [9] LAZAROS N, SIRAKOULIS G C, GASTERATOS A. Review of stereo vision algorithms: from software to hardware[J]. *International Journal of Optomechatronics*, 2008, 2(4): 435-462.
 - [10] BROWN M Z, BURSCHKA D, HAGER G D. Advances in computational stereo[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(8): 993-1008.
 - [11] SONG T, TANG B P, ZHAO M H, et al. An accurate 3-D fire location method based on sub-pixel edge detection and non-parametric stereo matching[J]. *Measurement*, 2014, 50(5): 160-171.
 - [12] HEO Y S, LEE K M, LEE S U. Joint depth map and color consistency estimation for stereo images with different illuminations and cameras[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(5): 1094-1106.
 - [13] SUNG M C, LEE S H, CHO N I. Stereo matching using multi-directional dynamic programming[C]//2006 International Symposium on Intelligent Signal Processing and Communications. New York: IEEE Press, 2006: 697-700.
 - [14] LI K P, WANG S N, YUAN M X, et al. Scale invariant control points based stereo matching for dynamic programming[C]//2009 9th International Conference on Electronic Measurement & Instruments. New York: IEEE Press, 2009: 769-774.
 - [15] LEUNG C, APPLETON B, SUN C M. Iterated dynamic programming and quadtree subregioning for fast stereo matching[J]. *Image and Vision Computing*, 2008, 26(10): 1371-1383.
 - [16] BLEYER M, GELAUTZ M. Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions[J]. *Signal Processing: Image Communication*, 2007, 22(2): 127-143.
 - [17] LEMPITSKY V, ROTHER C, BLAKE A. LogCut - efficient graph cut optimization for Markov random fields[C]//2007 IEEE 11th International Conference on Computer Vision. New York: IEEE Press, 2007: 1-8.
 - [18] WANG H Q, WU M, ZHANG Y B, et al. Effective stereo matching using reliable points based graph cut[C]//2013 Visual Communications and Image Processing (VCIP). New York: IEEE Press, 2013: 1-6.
 - [19] TOMBARI F, MATTOCCIA S, DI STEFANO L, et al. Classification and evaluation of cost aggregation methods for stereo correspondence[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2008: 1-8.
 - [20] YOON K J, KWEON I S. Locally adaptive support-weight approach for visual correspondence search[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2005: 924-931.
 - [21] HIRSCHMULLER H. Accurate and efficient stereo processing by semi-global matching and mutual information[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2005: 807-814.
 - [22] WANG Z F, ZHENG Z G. A region based stereo matching algorithm using cooperative optimization[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2008: 1-8.
 - [23] HOSNI A, BLEYER M, GELAUTZ M, et al. Local stereo matching using geodesic support weights[C]//2009 16th IEEE International Conference on Image Processing (ICIP). New York: IEEE Press, 2009: 2093-2096.
 - [24] YOON K J, JEONG Y, KWEON I S. Support aggregation via non-linear diffusion with disparity-dependent support-weights for stereo matching[M]//Computer Vision - ACCV 2009. Heidelberg: Springer, 2010: 25-36.
 - [25] ZHANG K, LU J B, YANG Q, et al. Real-time and accurate stereo: a scalable approach with bitwise fast voting on CUDA[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, 21(7): 867-878.
 - [26] DI H W, CHAI Y, LI K. A fast binocular vision stereo matching algorithm[J]. *Acta Optica Sinica*, 2009, 29(8): 2180-2194.
 - [27] MATSUO T, FUJITA S, FUKUSHIMA N, et al. Efficient edge-awareness propagation via single-map filtering for edge-preserving stereo matching[J]. *Proceedings of Spie the International Society for Optical Engineering*, 2015, 53(2): 79-91.
 - [28] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.
 - [29] ŽBONTAR J, LECUN Y. Computing the stereo matching cost with a convolutional neural network[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2015: 1592-1599.
 - [30] CHANG J, CHEN Y. Pyramid stereo matching network[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2018: 5410-5418.
 - [31] DUGGAL S, WANG S L, MA W C, et al. DeepPruner: learning efficient stereo matching via differentiable PatchMatch[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE Press, 2019: 4383-4392.
 - [32] TONIONI A, TOSI F, POGGI M, et al. Real-time self-adaptive deep stereo[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2019: 195-204.

- [33] GUO X Y, YANG K, YANG W K, et al. Group-wise correlation stereo network[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2019: 3268-3277.
- [34] ZHANG F H, PRISACARIU V, YANG R G, et al. GA-net: guided aggregation net for end-to-end stereo matching[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2019: 185-194.
- [35] BROMLEY J, BENTZ J W, BOTTOU L, et al. Signature verification using a “Siamese” time delay neural network[J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 1993, 7(4): 669-688.
- [36] ZAGORUYKO S, KOMODAKIS N. Learning to compare image patches via convolutional neural networks[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2015: 4353-4361.
- [37] KHAMIS S, FANELLO S, RHEMANN C, et al. StereoNet: guided hierarchical refinement for real-time edge-aware depth prediction[M]//Computer Vision – ECCV 2018. Heidelberg: Springer, 2018: 596-613.
- [38] LIU G D, JIANG G L, XIONG R, et al. Binocular depth estimation using convolutional neural network with Siamese branches[C]//2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). New York: IEEE Press, 2019: 1717-1722.
- [39] GUO C G, CHEN D Y, HUANG Z Q. Learning efficient stereo matching network with depth discontinuity aware super-resolution[J]. *IEEE Access*, 2019, 7: 159712-159723.
- [40] LUO J Y, XU Y, TANG C W, et al. Learning inverse mapping by AutoEncoder based generative adversarial nets[M]//Neural Information Processing. Heidelberg: Springer, 2017: 207-216.
- [41] PILZER A, XU D, PUSCAS M, et al. Unsupervised adversarial depth estimation using cycled generative networks[C]//2018 International Conference on 3D Vision (3DV). New York: IEEE Press, 2018: 587-595.
- [42] MATIAS L P N, SONS M, SOUZA J R, et al. VeIGAN: vectorial inpainting generative adversarial network for depth maps object removal[C]//2019 IEEE Intelligent Vehicles Symposium (IV). New York: IEEE Press, 2019: 310-316.
- [43] LORE K G, REDDY K, GIERING M, et al. Generative adversarial networks for depth map estimation from RGB video[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New York: IEEE Press, 2018: 1177-1185.
- [44] LIANG H, QI L, WANG S T, et al. Photometric stereo with only two images: a generative approach[C]//2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP). New York: IEEE Press, 2019: 363-368.
- [45] SCHARSTEIN D, SZELISKI R. High-accuracy stereo depth maps using structured light[C]//2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. New York: IEEE Press, 2003: 195-202.
- [46] SCHARSTEIN D, PAL C. Learning conditional random fields for stereo[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2007: 1-8.
- [47] HIRSCHMULLER H, SCHARSTEIN D. Evaluation of cost functions for stereo matching[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2007: 1-8.
- [48] SCHARSTEIN D, HIRSCHMÜLLER H, KITAJIMA Y, et al. High-resolution stereo datasets with subpixel-accurate ground truth[C]//German Conference on Pattern Recognition. New York: IEEE Press, 2014: 31-42.
- [49] 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.
- [50] MENZE M, GEIGER A. Object scene flow for autonomous vehicles[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2015: 3061-3070.
- [51] 李振雨, 王好臣. 基于视觉识别定位的苹果采摘系统研究[J]. *图学学报*, 2018, 39(3): 493-500.
LI Z Y, WANG H C. Research on apple picking system based on visual identification and location[J]. *Journal of Graphics*, 2018, 39(3): 493-500 (in Chinese).
- [52] BRANDAO P, MAZOMENOS E, STOYANOV D. Widening Siamese architectures for stereo matching[J]. *Pattern Recognition Letters*, 2019, 120: 75-81.