

基于 Stixel 的城市环境立体视觉场景表达

邹楠, 项志宇, 刘济林

(浙江大学信息与电子工程系, 浙江 杭州 310027)

摘要:正确的环境感知是自主车实现自主导航的重要前提。立体视觉在城市环境感知中具有重要的应用前景。相对于传统的基于像素视差的场景表达方式,棒状像素(Stixel)能够简化复杂的场景表达,提高后续处理速度,为进一步的场景理解应用提供有效支撑。介绍了 Stixel 构造的基本原理,并添加了 Stixel 视差计算过程,通过使用动态规划有效地提高了 Stixel 的视差精度。实验结果表明,改进后的 Stixel 算法具有准确率高,整体显示效果更平的特点,为后续物体跟踪和运动估计等应用提供了基础。

关键词:双目视觉;棒状像素;V 视差;动态规划

中图分类号:TP391.4

文献标识码:A

文章编号:1001-9146(2015)04-0031-05

0 引言

城市环境场景理解是交通监控和辅助驾驶的重要组成部分,具有重要的研究意义与应用价值。目前的场景理解研究主要集中在场景分割、三维恢复和对物体的识别等方面。根据所采用传感器的不同,环境感知手段主要有视觉、激光雷达、微波雷达以及多传感器信息融合等。其中双目视觉由于成本低廉,在研究中得到了更多的重视。一种基于双目视觉环境感知技术的三维场景理解技术—Stixel^[1-2](棒状像素)可以用来有效地表达城市的三维交通环境场景内容,并在视差图中分割出物体且获得良好的位置和尺寸估计,非常有利于基于视觉的辅助驾驶和导航。然而传统的 Stixel 直接以地面开始的视差值为 Stixel 的整体视差,容易受错误视差的影响,也没有考虑到左右视差棒之间的关系。本文提出采用动态规划技术,增加了 Stixel 视差计算过程,获得了更好构建效果。

1 Stixel 算法设计

1.1 流程设计

基于双目视觉的 Stixel 构建算法流程图如图 1 所示。

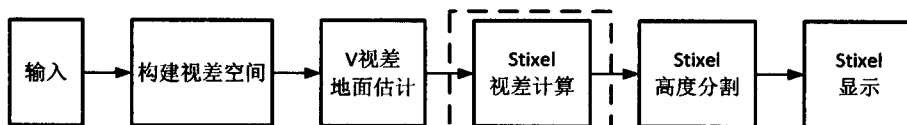


图 1 基于双目视觉的 Stixel 构建算法流程图

收稿日期:2014-11-03

基金项目:国家自然科学基金资助项目(61071219)

作者简介:邹楠(1991-),男,江苏南京人,在读博士生,计算机视觉. 通信作者:项志宇副教授, E-mail: xiangzy@zju.edu.cn.

图 1 中先将输入的原始双目图像数据进行立体匹配,构建视差空间,获得 Stixel 所需的视差数据。后面构建 Stixel 的 3 个主要部分:地面估计、Stixel 视差计算以及 Stixel 上高度分割。

1.2 构建视差空间

Stixel 构建算法建立在对视差概念的理解上,先对双目图像进行立体匹配。在本算法中,首先根据图像空间 (u, v) 及视差 d , 构建完备的视差代价空间 $c_m(u, v, d)$ 。并采取了半全局立体匹配算法 (Semi-Global stereo Matching, SGM)^[3]。这种方法基于局部最优的视差求取方法上,通过动态规划来提高计算结果平滑性,最终得到像素映射关系产生的视差图。

1.3 Stixel 构建算法

1.3.1 V 视差地面估计

V 视差算法的核心思想是通过沿水平方向,将相同的视差值进行累计。将原来的视差图重新构建成新的视差图^[4]。V 视差结果如图 2 所示,图 2(a)、图 2(b)为原始图像,图 2(c)是根据双目立体视觉原理,对左右图计算视差 $d = x_l - x_r = fb/z$ 所得图像。其中, x_l, x_r 分别表示对应点在左右两图在世界坐标系 (X, Y, Z) 中横坐标位置。 f 为相机焦距, b 为双目相机基线距离, z 为点到相机成像平面距离。同时假定世界坐标系中地面 $Z = pY + q$ 。根据相机标定原理,代入坐标公式,可以得到视差 $d = mV + n$ 。其中 m, n 都可以通过已知参数进行计算,即世界坐标系中地面可以映射成视差空间中的直线。之后可以通过 Hough 变换检测直线确定地面,如图 2(e) 所示。图 2(d) 为图 2(e) 内直线映射为视差图中的结果。

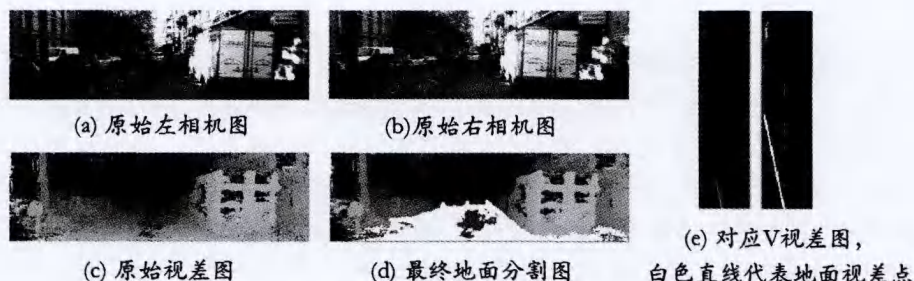


图 2 V 视差计算流程及结果

1.3.2 Stixel 视差计算

文献[1]构建 Stixel 的方案是基于地面边缘点的,以下边缘为基点建立 Stixel,以地面点为 Stixel 的视差值。这个方案容易受错误视差点的影响,而扩大了错误视差的传播。本文采用改进算法,使整个空间内的视差更为平滑、合理。具体方法在第 2 节中阐述。

1.3.3 Stixel 高度分割

1.3.2 节中得到了每一个横坐标所对应的 Stixel 表现形式,但在高度方向上没有确定边界。高度分割目的就是将物体的边缘信息明确地鉴定出来。通过设定隶属函数的方式,对 Stixel 上的每个像素确定其是否隶属于同一物体。1.3.2 节中得到了每一个横坐标 u 所对应的未进行高度分割的 Stixel 表现形式。再通过设定隶属函数的方式,对 Stixel 上的每个像素划分其是否隶属于同一物体。高度分割所追求的目标就是将物体的边缘信息明确地鉴定出来。假定隶属函数 $M_{u,v}(d)$, 当 $M_{u,v}(d)$ 为正时表示物体, $M_{u,v}(d)$ 为负时表示背景,定义函数如下:

$$M_{u,v}(d) = 2^{(1 - (\frac{d - \hat{d}_u}{\Delta D_u})^2)} - 1, \Delta D_u = \hat{d}_u - f_d(z_u + \Delta Z_u), f_d(z) = b \cdot f_x/z \quad (1)$$

式中, (u, \hat{d}_u) 对应于三维空间内坐标 (x_u, z_u) 。

对隶属函数 $M_{u,v}(d)$ 通过下式进行计算,获得明确的分割效果,当为边缘点时, $C(u, v)$ 取极值:

$$C(u, v) = \sum_{i=0}^{i=v-1} M_{u,v}(d(u, i)) - \sum_{i=v}^{i=v_f} M_{u,v}(d(u, i)) \quad (2)$$

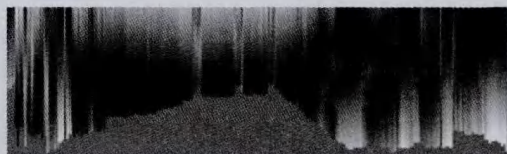
再对代价图像进行动态规划计算,并以代价函数 $C(u, v)$ 作为动态规划的数据项:

$$c_{u,v_0,v_1} = C(u, v_0) + S(u, v_0, v_1), S(u, v_0, v_1) = C_s |v_0 - v_1| \max\left(0, 1 - \frac{|z_u - z_{u+1}|}{N_z}\right) \quad (3)$$

式中, C_s 表示比例因子, $\max()$ 表示在深度不连续情况下的平滑度影响, $|z_u - z_{u+1}|$ 表示在不同列中的深度差 N_z 则为设定的深度参数。高度分割结果如图3所示。



(a)隶属函数及代价函数图像



(b)代价函数图像

图3 高度分割结果

2 Stixel 估计算法

本文采用动态规划的方法对 Stixel 进行估计。由前文的 SGM 立体匹配,已经从图像中获得了每个像素点所对应的局部最优的视差值。但是这样的视差值所对应的最优对象仅仅是像素本身,并不能对应于整体的图像。为了满足 Stixel 本身的定义,一根棒状像素,可以视为同一个元素,也就是说这个棒状像素拥有一致的视差值。这就需要在之前已经得到的地面的基础上对地面之上的像素值进行重新定义与规划。Stixel 视差为整个像素棒上的综合视差 d ,即左图棒 (u) 对应一个最优右图棒 ($u + d$)。为获取最优解,首先,改进方法对每列视差进行动态规划:

$$d_s^*(u) = \arg \min_{d(u)} \left(\sum_u c_s(u, d(u)) + \sum_{u_a, u_b} s_s(d(u_a), d(u_b)) \right), |u_a - u_b| = 1 \quad (4)$$

数据项根据 Stixel 上的假定的最小高度物体的综合亮度值进行计算:

$$c_s(u, d) = c_o(u, d) + c_g(u, d) \quad (5)$$

式中, $c_o(u, d)$ 和 $c_g(u, d)$ 分别表示物体和地面的匹配代价。 $c_o(u, d) = \sum_{v=v(h_o, d)}^{v(d)} c_m(u, v, d)$, $c_g(u, d) =$

$$\sum_{v=v(d)}^{|V|} c_m(u, v, f_{ground}(v))。$$

根据 V 视差地面估计,得到一条倾斜的曲线,即大致的地面视差 d 与纵坐标 v 的关系。 v 越大,对应的即是物体相对于摄像机平面越近,物体本身在左右两图之中的横向偏移量也就越大,也即是视差 d 越大。这条直线,可以直接定义为 v 与 d 的单一映射关系 $v(d)$ 。同样,这样的映射关系,根据世界坐标系中物体高度 h_o ,得到在对应视差 d 的高度 $v(h_o, d)$; $|V|$ 为行数。

平滑度估计如下:

$$s_s(d_a, d_b) = \begin{cases} \infty & d_a < d_b - 1 \\ c_o(u_a, d_a) & d_a = d_b - 1 \\ 0 & d_a > d_b - 1 \end{cases} \quad (6)$$

其中, d_a, d_b 表示当前 Stixel 与前一个的视差。在 $s_s = \infty$ 时保证不会有 Stixel 距离估计产生冲突;而当后列的深度略大于前列时,视为后列被遮挡,平滑项设定为物体代价,作为补偿,改进算法结果如图4所示。

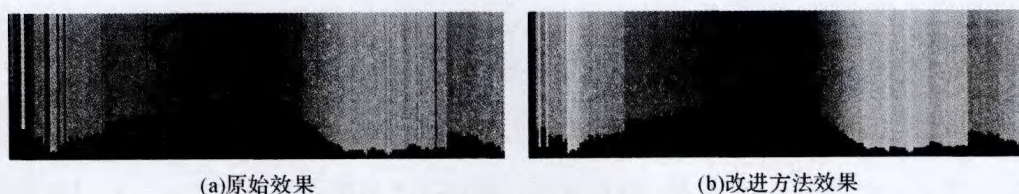


图4 改进算法前后视差计算效果

3 Stixel 构建结果

3.1 Stixel 定性显示效果

从图像中获得了每个 Stixel 的上下边缘,同时也获得了整体视差。通过投票法将多个连续列整合成较宽 Stixel,每个 Stixel 中都保存有其整体视差值。如图 5 所示。



图5 Stixel 显示效果

Stixel 显示也可以通过定义感兴趣区域(Region of Interest, ROI),并进行图像像素叠加、混合的方法来获得更好的显示效果。其中 Stixel 视差信息可以通过 RGB 色彩变化体现。

3.2 Stixel 构建结果定量分析

改进算法与原 Stixel 算法^[1]构建结果对比数据如表 1 所示。

表1 Stixel 视差显示结果对比表

样本 序列	总体参数		原 Stixel 算法		改进算法	
	图像大小	总 Stixel 数	正确 Stixel 数	正确率/%	正确 Stixel 数	正确率/%
1	456 × 375	76	65	85.53	72	94.74
2	642 × 375	107	88	82.24	102	95.33
3	1 014 × 375	169	145	85.80	160	94.67
4	1 242 × 375	207	176	85.02	191	92.27
平均值				84.79		93.92

从表 1 中可以看出,改进算法所得到的 Stixel 结果中,正确率均达到了 90% 以上的,高于原 Stixel 算法。可见改进后的算法获得的最终结果与实际结果相近,也就是说改进算法取到了较好效果。

总体来说城市场景可以获得较好的 Stixel 构建效果,但是由于其本身局限性,仍然存在一定误差。产生误差的因素主要有以下几点:1)整体的 Stixel 对立体匹配的效果要求较高。如果无法获得较好的立体匹配效果,则后续步骤的运算效果也会对应变差;2)由于 Stixel 的构建算法建立在 Stixel 垂直竖立在地面上这一假设。如果存在与假设不符的情况,会引起整个系统的误判。

4 结束语

本文根据双目立体视觉所构建的视差空间这一基本概念,研究了一种用棒状像素来替代普通像素对城市环境进行场景表达的方法。这种棒状像素可以用于表示城市场景之中地面、障碍物前景与背景之间分割。改进方案通过加入 Stixel 视差动态规划这一步骤,有效地提高了数据精度。总体上,本改进方法较为稳定,可以在多种环境场景之中进行构建及应用。

参考文献

- [1] Badino H, Franke U, Pfeiffer D. The stixel world-a compact medium level representation of the 3d-world[J]. Pattern Recognition. Springer Berlin Heidelberg, 2009: 51 - 60.
- [2] Benenson R, Timofte R, Van Gool L. Stixels estimation without depth map computation[C]//Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. Barcelona: IEEE, 2011: 2010 - 2017.
- [3] Hirschmüller H. Accurate and efficient stereo processing by semi-global matching and mutual information[J]. Proc Cvrp, 2005, 2(2): 807 - 814.
- [4] Labayrade R, Aubert D, Tarel J P. Real time obstacle detection in stereovision on non-flat road geometry through “v-disparity” representation[C]//Intelligent Vehicle Symposium, 2002. Versailles: IEEE, 2002, 2: 646 - 651.
- [5] Hirschmüller H. Semi-Global Matching-Motivation, Developments and Application[J]. Proc Photogrammetric Week, 2011: 173 - 184.

Stixel Based Urban Scene Representation by Stereo Vision

Zou Nan, Xiang Zhiyu, Liu Jilin

(Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou Zhejiang 310027, China)

Abstract: Accurate environmental perception is an important prerequisite for the autonomous navigation of autonomous vehicle. Stereo vision has important application prospect in the urban scene perception. In respect to the traditional way of representation of scenes based on pixel disparity, Stixel can simplify the representation of complex scenes, increase the processing speed and provide effective support for further scene understanding application. This paper introduces the basic principle of Stixel structure, and adds Stixel-disparity computation process improve the precision of Stixel-disparity by using dynamic programming. The experimental results show that the improved Stixel algorithm has high accuracy and the overall effects is smoother, provide a basis for the subsequent applications such as object tracking and motion estimation.

Key words: binocular; Stixel; V-disparity; dynamic programming