

一种新的快速立体视觉导航算法

邓 毅, 林学

(北京清华大学计算机科学与技术系, 普适计算教育部重点实验室, 北京 100084)

摘 要: 立体视觉技术被广泛的应用于视觉导航系统中. 由于要求计算快速, 以往的立体视觉导航系统通常采用局部匹配的方法, 其匹配精确度较差, 不能满足高性能视觉导航的需要. 为此, 我们提出了一种新的快速立体视觉匹配算法, 该算法将每一条外极线分割成亮度一致的线段, 并采用全局的树型动态规划进行优化求解, 既充分利用了全图的相关信息, 取得了较高的匹配准确度, 又减少了搜索变量的个数, 使得系统的效率得到大幅度的提高. 实验结果表明, 该算法能够取得与基于 MRF 的全局网格匹配算法相当的匹配结果, 而执行时间上却很快, 为视觉导航提供了可靠的环境三维数据.

关键词: 视觉导航; 立体匹配; 动态规划

中图分类号: TP242 **文献标识码:** A **文章编号:** 0372-2112 (2006) 11-2090-04

A Novel Fast Stereo Algorithm for Vision Navigation

DENG Yi, LIN Xue-yin

(Department of Computer Science and Technology, Key Lab of Pervasive Computing (MOE), Tsinghua University, Beijing 100084, China)

Abstract: Stereo technology has been widely used in visual navigation systems. Traditional stereo vision navigation systems usually use local method because of the real-time requirement. The accuracy of the results by local method, however, is very low and it can not be used in high performance vision navigation systems. In this paper we propose a novel fast stereo matching algorithm, which utilizes the segmentation information on the epipolar lines to form a tree and get optimized results by using tree dynamic programming procedure. The proposed algorithm can both use much of the relation information between neighboring pixels to achieve a more accurate matching result, and reduce the number of vertices to be solved so that the system becomes very efficient. Experimental results have shown that our algorithm can get good results comparable to the global MRF based algorithms, while the executing time is much shorter than the latter. Our algorithm is a fast algorithm which can provide reliable 3D information for vision navigation.

Key words: vision navigation; correspondence; dynamic programming

1 引言

立体视觉技术被广泛的应用于视觉导航、目标检测与跟踪、障碍物检测以及三维场景感知等场合, 但如何提高匹配准确度并能够进行快速计算, 仍然是一个比较困难的问题. 根据像素间联系信息的利用程度, 立体视觉匹配算法可以分为若干类型. 局部算法^[1]只利用像素周围邻近的局部信息进行像素块的匹配, 优点是快速, 但匹配精度较低. 网格算法通常把匹配问题用马尔科夫随机场 (MRF) 进行建模, 使得图中的所有像素通过网格方式连接起来. 由于充分利用了像素间的联系, 其匹配准确程度是所有方法中最好的. 然而, MRF 的优化通常是一个 NP 问题, 传统的模拟退火优化算法处理一幅图像需要若干小时或者更久的时间, 近几年出现的基于图分割 (Graph Cuts, GC)^[2]和可信度传递 (Belief Propagation, BP)^[3]的优化方法也需要数十秒到数分钟的处理时间, 这也使得它

们无法应用在视觉导航等要求计算速度快的场合. 外极线规划方法对单条外极线上的像素联合求解, 并利用动态规划进行优化. 在 Veksler^[2]的文章中, 提出了树型规划方法, 她把整幅图上的所有像素用一棵最小生成树连接, 采用树型动态规划算法进行优化求解. 文献 [2] 给出了当平滑函数为某些特定形式的时候, 复杂度可以降到 $O(nh)$ 的算法 (其中 n 是像素点的个数, h 是视差的级别数), 因此处理一幅 320×240 图像的时间约为若干分之一秒, 从而成为实时系统的一个较好的选择.

尽管 Veksler^[2]提出的树型规划算法的结果明显优于局部算法, 但是它的算法准确度还是逊于网格算法. 经过分析我们发现, 其主要原因在于它的动态规划算法比之于网格算法仍然丢掉了过多的像素间的联系信息. 图 1 绘制出了几种算法的等效边示意图. 设图像的大小为 $N \times N$, 网格算法 (图 1 (a)) 把所有的四连接网格边的信息都利用来进行统一求解,

这样的边的数目为 $2N(N-1) \approx 2N^2$; 基于外极线规划算法(图 1(b)), 则只考虑同一外极线内相邻像素的联系, 外极线间的信息被丢失, 其有效边(也就是保留下来的参与计算的边)的数目为 $N(N-1) \approx N^2$; 而 Veksler^[2] 的算法(图 1(c)), 为了区别于我们提出来的基于线段的算法, 我们称该算法为像素树型动态规划, P TDP, 尽管考虑了外极线内和外极线间的联系, 但由于需要用树来描述图像像素间的关系, 其有效边的数目为 $N^2 - 1 \approx N^2$. 由此可见, 以上两种动态规划算法, 与网格算法相比, 都丢失掉了约一半的有效边, 这是现有动态规划算法之所以不能取得与网格算法相当的准确度的主要原因.

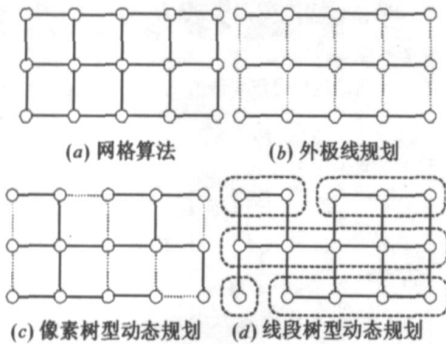


图 1 不同匹配算法的等效边示意图. 实线表示有效边, 虚线表示无效边(被丢弃的边). (d) 中的虚框表示同一段上的像素. 本图中, 相比网格算法, (b) (c) (d) 三个图中, 丢失掉的边的数目分别为 10, 8 和 2

为此, 我们考虑需要利用图像分割信息, 采用大粒度匹配使得我们能够保留更多的有效边, 进而提高匹配的准确度. 图像分割信息辅助立体视觉求解也是近年来的研究热点^[5-7]. 这些算法都利用了视差图中的跳变点, 只可能发生在分割块的边缘部分的假设, 提高算法在处理跳变点和无纹理区域的准确度. 然而, 这些算法都不是实时的算法, 其必须的区域分割本身就需要较长的时间, 因此无法满足视觉导航实时系统的要求.

我们设计了一种基于线段树的立体视觉匹配算法(LSPDP 算法). 我们首先用一个快速而有效的外极线分割算法, 把每一条外极线分割成若干条亮度一致的线段, 然后以这些线段作为结点建立搜索树, 并采用树型动态规划进行快速求解. 图 1(d) 中绘制了我们的线段树型动态规划算法的等效边情况. 从图中可见, 通过利用外极线亮度分割, 我们建立起的树结构能够保留像素间更多的联系信息, 相当于保留了更多的有效边, 最大程度的近似了全局网格, 从而可以提高匹配精确度. 由于线段的数量比像素的数量小, 搜索空间也大为减小, 这样就提高了算法的效率. 实验结果表明, 我们的算法输出的视差图, 其精度与网格算法相当, 而时间上却远小于网格算法, 对分辨率为 320×240 的图像进行处理, 单帧运行时间小于 100 毫秒, 可以满足视觉导航实时系统的要求.

2 基于线段的树型动态规划算法

2.1 线段分割

不失一般性, 我们假设立体相机的位置为标准配置, 也就

是相机的光轴互相平行, 且基线(光心的连线)与图像的水平方向一致. 此时水平扫描线即为外极线, 左右图像的对应点具有相同的纵坐标, 因此可以用对应点的水平坐标差值, 即视差 $d = x - x'$, 来表示图像中的对应关系. 令左右图像中的点的集合分别为 I_L 和 I_R , 并且选择左图作为参考图. 我们的算法是基于线段的匹配算法, 因此参考图上的每一条外极线首先被分割成亮度一致的线段, 记这些线段的集合为 S . 我们定义含相邻像素点的两条线段为相邻线段: 若 $s_1, s_2 \in S$ 为相邻线段, 则当且仅当存在 $p_1 \in s_1, p_2 \in s_2$, 而 p_1, p_2 为相邻像素点. 我们记所有相邻线段对的集合为 \mathcal{N} . 相邻线段的公共边界长度, 定义为相邻的像素点对的个数, 记为 $L_c(s_1, s_2)$. 如果 s_1, s_2 处于同一外极线上(图 2(a)), $L_c(s_1, s_2) = 1$; 否则, s_1, s_2 处于相邻的外极线上, 此时 $L_c(s_1, s_2) = \min\{x'_1, x'_2\} - \max\{x''_1, x''_2\} + 1$, 其中 x''_1, x''_2, x'_1, x'_2 分别为 s_1, s_2 的左右端点的横坐标.

2.2 搜索树建立算法

我们采用图论中的图结构来描述分割后各线段间的关系, 令 $G(V, E)$ 为一棵由结点集和边集组成的图. 节点集 V 由外极线分割成的线段组成(在这里就是 S). 边集 E 中的元素则表示了线段间的相邻关系 \mathcal{N} . 通常情况下, G 是一个存在大量环的图, 此时匹配算法的求解需要采用 BP 或者图分割的方法, 效率较低. 因此, 考虑找到 G 的一个生成树 G^T 来近似 G 中的相邻关系. 为了保证 G^T 能够更好的近似 G , 保留尽量多的相关信息, 对于生成树的选择, 我们规定了如下两个原则:

- (1) 在 G^T 中相邻的线段, 应该有相近的视差, 他们应该尽可能的属于空间中的同一个区域, 因而也更需要平滑约束. 此原则与文献[4]中的原则相似, 我们同时假设相似亮度的相邻像素应该有相近的视差;
- (2) 相邻的线段应该有尽量长的公共边界, 从而使得 G^T 能够更大程度地近似 G .

结合以上两个原则, 我们对于两个相邻线段 s 和 t 设计了如下的权值 w_{st} 的计算公式:

$$w_{st} = I_{\max} - \sigma(\bar{I}_s, \bar{I}_t) \cdot L_c(s, t) \quad (1)$$

L_{\max} 是所有线段长度的最大值, \bar{I}_s 和 \bar{I}_t 分别为线段 s 和 t 上像素的平均亮度, $\sigma(\bar{I}_s, \bar{I}_t)$ 是取值为 $[0, 1]$ 的亮度相似度函数(0 表示完全不相似, 1 表示非常相似或相同). 采用此权值函数对 G 中的每条边计算权值, 并利用通用的最小生成树(MST)算法, 就可以在 $O(|S|)$ 的时间内建立起所需要的搜索树.

2.3 基于线段的立体视觉匹配

立体视觉匹配的目标是计算得到每个线段 $s \in S$ 的视差 d_s , 为此我们通过优化如下能量函数求解视差图:

$$E(D) = \sum_{s \in S} E_{\text{data}}^s(d_s) + \sum_{s, t \in S} E_{\text{smooth}}^{s,t}(d_s, d_t) \quad (2)$$

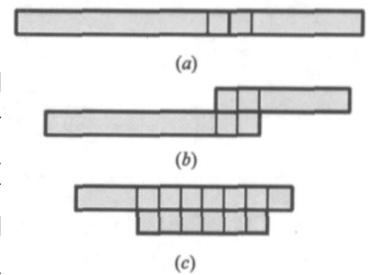


图 2 几种相邻线段的配置. 图中标记出来的像素点为相邻像素点. (a) 中的相邻线段处于同一外极线上, (b) 和 (c) 中的相邻线段处于相邻外极线上

其中 $N^T \subset \mathcal{N}$ 为搜索树保留的邻域线段对的集合, $E_{\text{data}}^s(d_s)$ 为线段 s 在给定视差 d_s 时的匹配误差能量, 其形式可以根据需要定义. 在我们的实验中, $E_{\text{data}}^s(d_s)$ 取为线段上各点与其对应点的亮度差异绝对值之和. $E_{\text{smooth}}^{s,l}(d_s, d_l)$ 为相邻线段间的平滑约束, 为了防止模糊物体边界而又能够更好的匹配图像垂直方向倾斜的平面, 我们采用了如下的平滑函数:

$$E_{\text{smooth}}^{s,l}(d_s, d_l) = v_{sl} \cdot L_c(s, l) \cdot \begin{cases} 0, & d_s = d_l \\ \tau_1, & |d_s - d_l| = 1, \\ \tau_2, & |d_s - d_l| > 1 \end{cases} \quad (3)$$

其中 v_{sl} 控制平滑能量的强度, τ_1 和 τ_2 为两个常数.

通过树型动态规划算法优化式(2)的能量函数, 我们就可以求解出最优的视差图.

3 算法实现

图 3 为 LSTDP 算法的总体流程图, 下面分别介绍各模块的功能与实现.

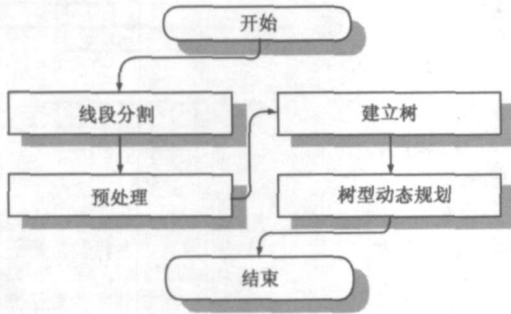


图 3 LSTDP 算法总体流程图

3.1 线段分割模块

线段分割模块负责将参考图上的每一条外极线分割为亮度相近的像素组成的线段. 考虑到效率的原因, 我们没有选择诸如均值移动 (Mean Shift)^[8] 或者图分割 (Normalized Cuts)^[9] 的算法, 这些算法本身有可能成为整个算法的瓶颈. 为此, 我们设计了一个简单而有效的快速分割算法.

我们的算法分为三步:

(1) 初始标记

从左到右扫描外极线上的每一个像素, 设定两个寄存器 V_{max} 和 V_{min} 保存当前亮度 (各分量) 的最大和最小值, 如果 $V_{\text{max}} - V_{\text{min}}$ 大于某个阈值 T_{seg} , 就在当前位置设立标记, 同时重置寄存器. 两个相继标记间的线段, 认为是一个分割出的线段.

(2) 精调标记位置

步骤(1)中分割出来的标记, 往往不是精确处于边缘点位置. 因此, 我们调整他们的位置到相邻的水平梯度极大值点上. 设调整前后标记点的水平坐标分别为 x 和 x' , 调整的方法为 $x' = \arg \max_{x-5 \leq x' \leq x+5} |I(x') -$

$I(x' - 1)|$, 同时我们不改变标记点间的顺序关系.

(3) 删除孤立点

实验中, 我们发现, 由于噪声的原因, 会分割出来少量的孤立点, 影响匹配, 因此需要检查每个标记点周围的小领域 (实验中为该点周围 5×5 的区域), 如果没有别的标记点, 则将该孤立点删除.

图 4 为我们设计的线段分割算法对某幅图像进行分割的中间和最后结果. 实验结果表明, 这个分割算法快速而有效.

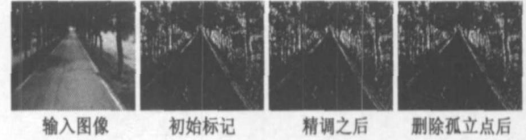


图 4 快速线段分割算法的中间和最后结果

3.2 预处理模块

在实际的系统中, 由于硬件的不对称性或者由于自动增益系统的调整, 造成输入的左右图像的灰度分布发生偏差, 对应点的亮度差异较大, 影响匹配. 因此, 在匹配前的预处理模块中采用 Cox 等人^[10] 提出的动态直方图调整 (DHW, Dynamic Histogram Warping) 算法进行直方图调整, 使得两幅图的亮度分布基本一致.

3.3 建立树模块

这个模块中, 我们利用式(1)计算相邻线段间的权值, 然后建立最小生成树, 从而形成相邻线段对集合 \mathcal{N}^T .

4 树型动态规划模块

最后利用树型动态规划算法优化公式(1)中的能量函数. 由于篇幅所限, 动态规划的算法细节不在这里列出, 感兴趣的读者请参考文献[4]和文献[11].

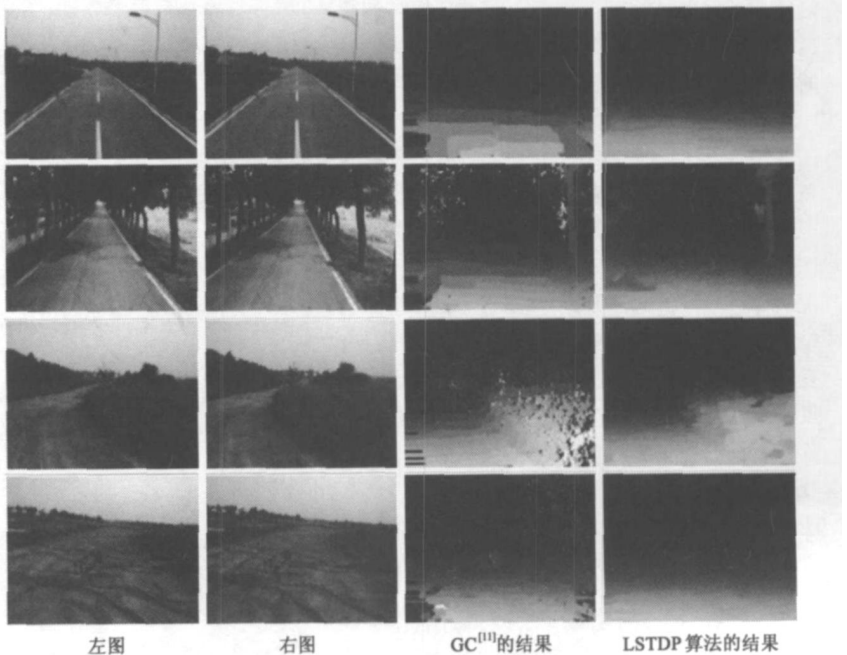


图 5 实验结果图. 上两行为道路图像处理结果图, 下两行为越野地形图像的结果图

5 实验及结果

我们采用 Birthfield 和 Tomasi^[12] 的插值算法计算公式(1)中的匹配能量函数,以减小图像采样对匹配的影响.公式(3)中的强度系数 v_{st} 按照下式进行计算:

$$v_{st} = c_1 + c_2 \cdot \sigma(\bar{I}_s, \bar{I}_t).$$

上式使得相似颜色的相邻线段更加平滑.

我们在一个视觉导航的实时系统中测试我们的算法.硬件条件为:实验机器 Intel P4 2.4G, 相机分辨率为 320×240. 算法的匹配结果显示于图5中.作为对比,我们使用 GC 算法(文献 2)) 作者网站上的代码,对亮度校正后的图像进行处理.从实验结果中可以看出,我们的算法给出的结果的准确度与网格算法相当甚至更好.在相同机器条件下,GC 算法的执行时间约为 1 分钟,而 LSTD 算法的单帧运行时间小于 60 毫秒.可见我们的算法匹配速度快,准确度也较高.

6 结论

本文提出了一个快速的立体视觉匹配算法.通过利用分割信息及树型动态规划,该算法能够最大限度的利用像素间的相关信息,而又保持了动态规划的高效率.算法在实验的实时系统上,表现出了较高的性能和极高的速度.

我们今后的改进方向将集中在如何更好的提高性能方面,包括对算法进行并行化,进一步提高匹配速度、更好的处理遮挡问题、使用新的约束假设等.

参考文献:

- [1] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. International Journal on Computer Vision, 2002, 47(1/2/3): 7–42.
- [2] Kolmogorov V, Zabih R. Computing visual correspondence with occlusions using graph cuts [A]. IEEE International Conference on Computer Vision [C]. Vancouver: IEEE Computer Society, 2001. 508–512.
- [3] Sun J, Li Y, Kang S B, Shum H Y. Symmetric stereo matching for occlusion handling [A]. IEEE International Conference on Computer Vision and Pattern Recognition [C]. San Diego: IEEE Computer Society, 2005. 556–571.
- [4] Veksler O. Stereo correspondence by dynamic programming on a tree [A]. IEEE International Conference on Computer Vision and Pattern Recognition [C]. San Diego: IEEE Computer Society, 2005. 384–390.
- [5] Tao H, Sawhney H S, Kumar R. A global matching framework for stereo computation [A]. IEEE International Conference on Computer Vision [C]. Vancouver: IEEE Computer Society,

2001. 532–539.

- [6] Hong L, Chen G. Segment based stereo matching using graph cuts [A]. IEEE International Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE Computer Society, 2004. 74–81.
- [7] Wei Y, Quan L. Region based progressive stereo matching [A]. IEEE International Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE Computer Society, 2004. 106–113.
- [8] Comaniciu D, Meer P. Robust analysis of feature spaces: Color image segmentation [A]. IEEE International Conference on Computer Vision and Pattern Recognition [C]. San Juan: IEEE Computer Society, 1997. 750–755.
- [9] Shi J, Malik J. Normalized cuts and image segmentation [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888–905.
- [10] Cox I J, Roy S, Hingorani S L. Dynamic histogram warping of image pairs for constant image brightness [A]. IEEE International Conference on Image Processing [C]. Washington: IEEE Computer Society, 1995. 366–369.
- [11] Felzenszwalb P, Huttenlocher D. Pictorial structures for object recognition [J]. International Journal on Computer Vision, 2005, 61(1): 55–79.
- [12] Birthfield S, Tomasi C. A pixel dissimilarity measure that is insensitive to image sampling [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1998, 20(4): 401–406.

作者简介:



邓 毅 男, 广东阳山人, 1978 年生于广西柳州, 本科就读于清华大学计算机科学与技术系, 现在普适计算国家重点实验室就读博士研究生, 主要研究领域为计算机视觉、数字图像处理、模式识别以及立体视觉匹配等.

E-mail: dengyi00@mails.tsinghua.edu.cn



林 学 男, 1940 年生于浙江瑞安, 清华大学计算机科学与技术系教授, 博士生导师, 主要研究领域为计算机视觉、模式识别.