

文章编号:1001-9081(2009)10-2695-03

二阶先验模型下 SA-GA 混合优化策略的图像景深提取

杨竣辉¹, 黄 婵^{1,2}

(1. 江西理工大学 信息工程学院, 江西 赣州 341000; 2. 赣南教育学院 计算机系, 江西 赣州 341000)

(jwejhy@126.com)

摘 要: 解决从不同角度拍摄的一系列平面图像中解算图像中物体的景深问题, 使用一个新颖的二阶先验模型来建模光滑 3D 物体的表面信息, 此模型比传统的一阶模型有很大优势。同时提出一个混合优化策略——模拟退火算法(SA)和遗传算法(GA)来提高优化效率。实验结果表明: 二阶先验模型在解算典型场景的景深时要优于一阶模型。混合优化比单独的 SA、GA 优化效果好。

关键词: 图像 3D 重建; 二阶先验模型; 建模; 模拟退火算法; 遗传算法

中图分类号: TP391.41 **文献标志码:** A

Hybrid optimization of SA-GA depth map extraction base on second-order prior model

YANG Jun-hui¹, HUANG Chan^{1,2}

(1. School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou Jiangxi 341000, China;

2. Department of Computer, Gangnan Institute of Education, Ganzhou Jiangxi 341000, China)

Abstract: This work deals with the depth information estimation of 3-D surface in a series of images from different shot angle. The proposed method used a new second-order prior model to model the smooth 3D surface information which cause intractable (non-submodular) optimization problems; improved optimization efficiency by using the hybrid optimization strategy of Simulated Annealing (SA) algorithm and Genetic Algorithm (GA). Experimental results demonstrate that the second-order prior model is a better model of typical scenes than the first-order prior one and the performance of the hybrid algorithm outperforms SA and GA alone.

Key words: Image Based Rendering (IBR); second-order prior model; modeling; Simulated Annealing (SA) algorithm; Genetic Algorithm (GA)

0 引言

基于图像进行 3D 重建 (Image Based Rendering, IBR) 的关键技术是从二个以上不同角度拍摄的同一场景的图像中提取三维目标的表面深度信息。事实上, 物体的三维几何信息编码在物体的不同角度看到的图像中, 自然界数百万年的进化使得动物的大脑仍以一种不为人知的神经机制从双目视觉中解算所看到的物体的三维几何信息^[1]。最近几年来基于图像的物体深度信息提取已成为一个新兴的研究领域, 各种算法不断涌出^[2]。这些方法中, 基于马尔科夫随机场 (Markov Random Field, MRF) 的重建成为一个热门的研究方向^[2-4]。通过在贝叶斯后验概率公式中引入平滑先验模型 (MAP-MRF), 将 IBR 转化为能量函数最小化问题, 其根本就是算法问题。目前, 能量函数的全局最小值算法较多, 具有代表性的有图切算法^[4], 但是它的局限是能量函数的平滑分量只能使用一阶先验模型。

截止目前为止, 大部分基于 MRF 的重建算法都是使用一阶先验模型, 而一阶模型只能对比较光滑、棱角起伏变化不大的物体表面建模, 对于表面棱角起伏变化剧烈的物体表面的效果并不好, 显然后一种物体更符合自然界中物体的天然形状。最近也有报告^[5]称一阶先验模型在高曲率表面产生的结果不精确, 如图 1。在使用“teddy”系列图像^[6]的实验中, 图 1(c) 左边的星形物体就因为一阶模型偏好起伏不大表面

的特性而与长方形的物体有了同样的深度值。

为解决一阶模型的不精确问题, 提出用一个二阶模型来建模物体的表面深度信息。相应地, 为解决的传统优化算法对二阶模型效率不高的问题, 给出一个新颖的 SA-GA 混合优化策略来最小化能量函数, 并在 Middlebury^[8]提供的测试集上检验性能。



图 1 一阶先验模型在高曲率表面产生的结果

1 二阶先验模型数学模型

对应 IBR 问题, 要做的是从给定的 $N+1$ 的图像集 $\{I_i\}_{i=0}^N$ 的每个像素中解算深度图 D 。因此可在贝叶斯后验模型中, 通过后验概率 $P(D | \{I_i\}_{i=0}^N)$ 来解算深度信息 D 。

$$P(D | \{I_i\}_{i=0}^N) = \frac{p(\{I_i\}_{i=0}^N | D)p(D)}{p(\{I_i\}_{i=0}^N)} \quad (1)$$

分子上的第一个元素就是后验概率, 有时候它也被称作图像数据分量, 它用来衡量 D 符合观测到图像的程度。分子上的第二个元素经常被称作先验平滑模型, 在观测到图像之前, 它用来度量一个 D 实例是先验概率的可能性。

为了找到一个最大可能性的深度图 D , 需最大化一个后

收稿日期: 2009-04-07。 基金项目: 江西省教育厅科技项目 (GJJ08286)。

作者简介: 杨竣辉 (1981-), 男, 江西瑞金人, 讲师, 硕士研究生, 主要研究方向: 图像处理、模式识别、数据交换; 黄婵 (1978-), 女, 江西赣州人, 讲师, 硕士研究生, 主要研究方向: 数据挖掘。

验概率来得到 D , 即公式 $D = \operatorname{argmax}_D p(D | \{II\}_{i=0}^N)$ 。根据 Hamersley-Clifford 理论^[3], MRF 可等价于吉布斯随机场, 从而有了能量函数:

$$E[D] = -\log(P(\{II\}_{i=0}^N | D)P(D)) = E_D + E_P \quad (2)$$

这里:

$$E_D = -\log P(\{II\}_{i=0}^N | D) \quad (3)$$

$$E_P = -\log P(D) \quad (4)$$

通过用数据分量来描述 E_D , 用先验分量来描述 E_P , 从而最大化一个后验概率就转化为上述的能量函数:

$$\hat{D} = \operatorname{argmax}_D E[D] = E_D \quad (5)$$

要被最小化的能量函数是一个以深度图 D 为自变量的函数, 它由两个分量组成: 几何可见性推理和光滑元素。通过定义一个二维向量 \mathbf{x} , 表示图像里像素的位置, 像素元素被写作 $I_0(\mathbf{x})$, 其相应的深度信息写作 $D(\mathbf{x})$ 。同时还定义了一个映射函数 $\{\pi_i(\mathbf{x}, d): \mathbf{R}^2 \rightarrow \mathbf{R}^2\}_{i=1}^N$, 其中 $\pi_i(\mathbf{x}, d)$ 是从 3D 点深度信息到图像 i 的像素 \mathbf{x} 的映射。对于修正后的图像序列来说, 只需要两幅图像其中 $N = 1$ 和 π_1 。这样就有了一个简明的定义 $\pi_i(\mathbf{x}, d) = \mathbf{x} + [d, 0]$ 。于是可使用简化形式 $I_i^r(\mathbf{x}, d) = I_i[\pi_i(\mathbf{x}, d)]$ 来表示当 \mathbf{x} 的深度为 d 时图像 i 中像素 \mathbf{x} 的颜色。

1.1 数据分量

采用标准的一致性图像数据分量, 它有着以下的形式:

$$E_d(D) = \sum_{\mathbf{x}} \sum_{i=1}^N f(I_i^r(\mathbf{x}, D(\mathbf{x})) - I_0); V_x^i \quad (6)$$

这里 V_x^i 是一个可见性标志, 或可写为 $V_i(\mathbf{x}, D)$ ^[4], 说明 $(\mathbf{x}, D(\mathbf{x}))$ 定义的 3D 点在图像 i 中是否可见。 V_x^i 一致性矩阵 f 定义为:

$$f(\Delta I, V) = \begin{cases} \rho_d(\Delta I), & V = 1 \\ v, & V = 0 \end{cases} \quad (7)$$

这里的 V 是被遮挡像素的惩罚代价值, ρ_d 是对颜色差异的衡量, 定义为:

$$\rho_d(I) = -\log(1 + e^{(-\|I\|^{2/\sigma_d})}) \quad (8)$$

这里 σ_d 是图像中的噪声水平。

1.2 先验模型分量

在像素邻域 θ 上, 平滑先验定义一个代价函数 $\rho_s(\cdot)$ 。 E_p 是在定义好的邻域 N 上的所有代价函数的总和, 故:

$$E_p(D) = \sum_{N \in \theta} \rho_s(S(N, D)) \quad (9)$$

这里 $\rho_s(S) = \operatorname{Min}(n(\sigma_s, |s|))$ 是截断性的线性核心。

二阶先验模型是定义在三阶邻域上用来估算深度图的二阶微分:

$$S(\{p, q, r\}, D) = D(p) - 2D(q) + D(r) \quad (10)$$

这里的邻域 $N = \{p, q, r\}$ 来自于输入图像集的 3×1 和 1×3 小块。相比使用传统的一阶先验模型 $S(\{p, q\}, D) = D(p) - D(q)$, 二阶模型的这个函数单调递增。

2 混合优化策略

这部分将阐述如何最小化能量函数。SA 有局部搜索能力强, 但收敛速度慢的特点; GA 有搜寻空间大, 但局部寻优能力弱的特点^[9]。根据它们各自特点, 自然会想到通过结合 SA 和 GA 各自优点从而快速得到能量函数的全局最小值。在这种混合优化策略下, 通过扩展适应度函数即能量函数 $E[D]$, 从而避免在使用轮盘赌方式的演化早期出现的早熟问题, 采用下面的适应度扩展方法:

$$fie_i = \frac{e^{f_i/T}}{\sum_{i=1}^M e^{f_i/T}} \quad (11)$$

在这 f_i 是第 i 个染色体的适应度值。交叉率和变异率随着演变系数的增加呈指数下降, 也就是:

$$\begin{cases} P_{\text{cross}} = e^{(-(n-1)/(Gen-1))} P_{\text{cross}} \\ P_{\text{mute}} = e^{(-(n-1)/(Gen-1))} P_{\text{mute}} \end{cases} \quad (12)$$

这种现象背后的原因是第一代的交叉率和变异率在变大, SA-GA 就能得到更大的解决方案空间, 以便 SA-GA 更快地收敛。当交叉率和变异率在最后一代变小时, SA-GA 能得到更好的搜索能力并且能得到更低的能量函数值。在编码阶段通过使用二维染色体编码, 在种群的初始化阶段, 为了保持个体的多样性, 加上一个随机的变化量。

$$T_i(i, j) = T_i(i, j) - rand \times T_i(i, j) \times k \quad (13)$$

在这 k 是小摇摆值, $rand$ 是随机数。

在基因的重组和变异操作时, 二阶 MRF 势团和二维窗口被整合在一个个体中。实验中使用的 SA 温度公式为:

$$T = \frac{C_0}{\ln(C_1 + n)} \quad (14)$$

在这 C_0 和 C_1 是常数, n 是第 n 次迭代。

因此, SA-GA 杂交优化可总结如下:

- 1) 初始化参量, 包括种群数目、交叉率、变异率、初始温度参量等;
- 2) 使用式(13) 初始化种群;
- 3) 对于每个个体, 逐像素扫描深度图, 随机产生新的深度值, 并根据 $E[D]$ 的定义计算能量值;
- 4) 如果 $E[D] > 1$, 则用新深度值替换原值, 否则用一个 0 和 1 范围的随机数替换;
- 5) 计算扩展形式的适应度函数;
- 6) 基于赌轮盘规则做选择;
- 7) 使用式(12) 计算交叉率和变异率;
- 8) 用 2 维窗口重组, 并用 2 维窗口实施变异;
- 9) 使用式(14) 降低温度;
- 10) 如果温度不够低, 返回第 3), 否则终止。

算法的时间复杂是:

$$T(n) = SA_n \times Ps \times C_n \times N_n \times k \times n = O(n)$$

其中: SA_n, Ps, C_n, N_n, k 分别是代表迭代次数、种群大小、分类个数、邻域数目和独立于 n 的计算系数; n 表示图像的像素个数。

3 实验

在此部分, 通过实验对比一阶和二阶先验模型以及混合优化和 GA, SA 优化的效果。实验中的参数经验地选择如下: 群体大小为 20, 演变代数 50, 交叉率为 0.1, 变异率为 0.01, $C_0 = 1.0, C_1 = 2.0$, 最终结束温度时间 $T_e = 0.18$ 。

3.1 一阶和二阶对比

使用“teddy”系列图像^[6]来比较一阶二阶深度图结果的精确性。图 2 显示了它们各自的结果与 Middlebury 大学提供的测试集的标准深度图的比较。从图 2 中可以很容易看出, 二阶的错误百分比下降得非常快, 而且在迭代的第 50 次错误百分比接近于标准深度图; 也可以看到二阶的错误百分比在每一个迭代阶段都优于一阶。

从实验可以得出, 与一阶相比二阶的优越性在于它可以很好地对曲面变化剧烈的表面进行先验建模。

3.2 混合优化的性能

接下来, 再来比较混合策略和 SA-GA 的性能。实验图像序列使用库中的“baby”, 图 3 显示提出的混合优化不仅减少

了迭代时间而且能达到最低的能量函数值。混合优化收敛得最快,得到的能量值最低,GA 在大概 15 代时陷入了局部最小而不能自拔。SA 比 GA 得到的能量值低,但是却需要很长的时间来达到这个效果。混合优化解决了不成熟收敛、收敛时间过长、容易陷入局部最小而不能得到全局最优结果的问题。

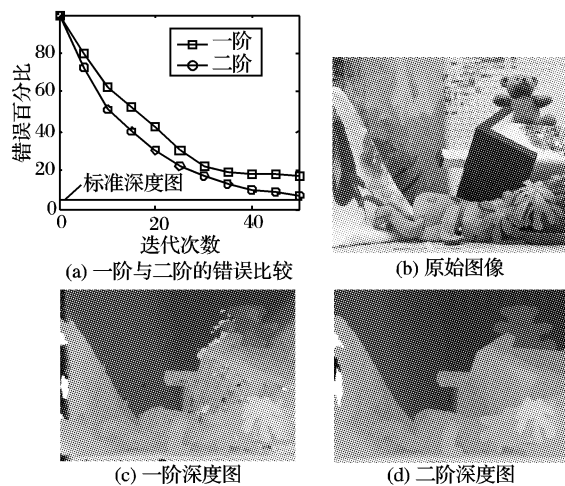


图 2 一阶与二阶的比较

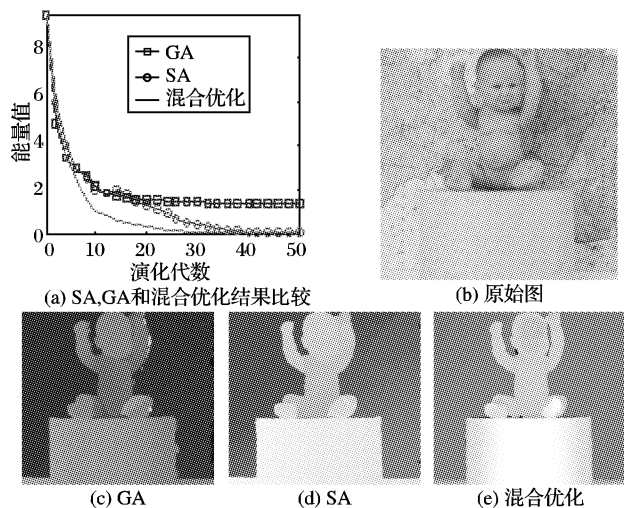


图 3 三种算法的实验结果

虽然,理论上 SA 可以找到全局最优值,但是计算效率和性能直接存在冲突,因为马尔科夫链控制的不易,很难判断是否 SA 在每一个温度阶段能达到平衡点。SA 虽然有着不错的“爬坡”性能,但是却在消耗时间上差强人意。与 SA 相反,

GA 的搜索空间很大,但其爬坡性能很差。

4 结语

本文对二阶先验模型可以整合基于图像的深度信息重建的数学模型进行解算,而且有着优于传统一阶模型的高精度结果。为了高效地优化这种能量函数,提出 SA-GA 混合优化策略。实验结果显示混合优化有着优于单独使用 SA 或 GA 的性能,可以高精度、快速地得到深度图并且在全局和局部都有着很强的搜索能力,它在很短的时间可以找到所有的极值,而且不要求很严格的初始化参数。

参考文献:

- [1] CHANG N, LIU T-M, TSAI T-H, *et al.* Real-time DSP implementation on local stereo matching[C]// Proceedings of 2007 IEEE International Conference on Multimedia and Expo. [S. l.]: IEEE Press, 2007: 2090 - 2093.
- [2] GHERARDI R. Confidence-based cost modulation for stereo matching[C/OL]. [2009 - 01 - 01]. <http://figment.csee.usf.edu/~sfefilat/data/papers/MoBT8.39.pdf>.
- [3] TAPPEN M F, FREEMAN W T. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters[C]// ICCV: Proceedings of the Ninth IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2003, 2: 900 - 907.
- [4] WEI YI-CHEN, QUAN LONG. Asymmetrical occlusion handling using graph cut for multi-view stereo[C]// CVPR: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2005, 2: 902 - 909.
- [5] ISHIKAWA H, GEIGER D. Rethinking the prior model for stereo[C]// ECCV: Proceedings of 9th European Conference on Computer Vision, LNCS 3953. Heidelberg: Springer-Verlag, 2006: 526 - 537.
- [6] SCHARSTEIN D, SZELISKI R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. International Journal of Computer Vision. 2002, 47(1/3): 7 - 42.
- [7] ZITNICK C L, KANADE T. A cooperative algorithm for stereo matching and occlusion detection[J]. IEEE TPAMI, 2000, 22(7): 675 - 684.
- [8] Daniel scharstein[EB/OL]. [2009 - 01 - 01]. <http://www.middlebury.edu/about/pubaff/facultyexperts/scharstein.htm>
- [9] 邵平凡, 万程鹏. 求解全局优化问题的遗传退火算法[J]. 计算机工程与应用, 2007, 43(12): 62 - 65.

(上接第 2694 页)

- [3] BERGN J R, ANANDAN P, HANNA K J, *et al.* Hierarchical model-based motion estimation[C]// Proceedings of the Second European Conference on Computer Vision, LNCS 588. London: Springer-Verlag, 1992: 237 - 252.
- [4] BROWN L G. A survey of image registration techniques[J]. ACM Computing Survey, 1994, 24(4): 325 - 367.
- [5] BURT P J. Fast filter transform for image processing[J]. Computer Graphics and Image Processing, 1981, 1(6): 20 - 51.
- [6] EFROS A, FREEMAN W. Image quilting for texture synthesis and transfer[C]// Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 2001: 341 - 346.
- [7] PEREZ P, GANGNET M, BLAKE A. Poisson image editing[J].

ACM Transactions on Graphics, 2003, 22(3): 313 - 318.

- [8] BURT P J, ADELSON E H. A multiresolution spline with application to image mosaics[J]. ACM Transactions on Graphics, 1983, 2(4): 217 - 236.
- [9] SZELISKI R. Video mosaics for virtual environments[J]. IEEE Computer Graphics and Applications, 1996, 16(2): 22 - 30.
- [10] YAMASHITA N, FUKUSHIMA M. On the rate of convergence of the Levenberg-Marquardt method[J]. Computing, 2001, 15: 239 - 249.
- [11] LEVENBERG K. A method for the solution of certain nonlinear problems in least squares[J]. Quarterly Journal of Applied Mathematics, 1944, 2(2): 164 - 166.
- [12] 杨柳, 陈燕萍. 求解非线性方程组的一种新的全局收敛 Levenberg-Marquardt 算法[J]. 计算数学, 2008, 30(4): 389 - 395.