

# 基于自适应分水岭的图割的立体匹配算法

祝世平 杨 柳

(北京航空航天大学仪器科学与光电工程学院测控与信息技术系, 北京 100191)

**摘要** 基于对现有图割算法的研究, 设计了基于自适应分水岭算法并使用非参数深度平滑模型来建立图割能量方程的立体匹配方法。提出了新的自适应局部阈值方法, 并将其应用于分水岭结合 Prim 算法的区域融合中。该方法选取相同亮度的像素作为同一个特征矢量形成像素组层, 这样两幅或多幅图像的匹配可以在特征区域像素组层来计算, 大大减少了数据量。在最小化能量方程时, 基于像素组层优化现有的  $\alpha$ -扩展算法, 降低运行时间。通过 Middlebury 测试平台对算法定量评估得出在所有区域的误匹配率、非遮挡区域以及深度不连续区域的误匹配率都控制在 8.5% 以内, 在 Middlebury 测试平台 135 组数据中排名第 19 位。

**关键词** 机器视觉; 自适应分水岭; 图割; 能量函数; 立体匹配

**中图分类号** TN919.82 **文献标识码** A **doi**: 10.3788/AOS201333.0315004

## Stereo Matching Algorithm with Graph Cuts Based on Adaptive Watershed

Zhu Shiping Yang Liu

(Department of Measurement Control and Information Technology, School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China)

**Abstract** Based on the existing graph cut algorithms, a stereo matching method which is based on adaptive watershed algorithm and non-parametric depth smoothing model to create an energy equation of graph cut is designed. A novel adaptive local threshold method is proposed, which is applied to region integration by combining the watershed with Prim's algorithm. The proposed method selects the pixels with same brightness as a feature vector and forms pixels group, thus two or more images can be matched in the pixels group layer of feature region, then the amount of data can be greatly reduced. When the energy equation is minimized, pixels groups layer can be used to optimize the  $\alpha$ -expansion algorithm, and reduce the running time. Experimental results indicate that the error match ratios of all regions, non-occluded area and depth discontinuity regions are all less than 8.5%. The proposed algorithm is in the place of 19 among all the 135 algorithms in Middlebury testing platform.

**Key words** machine vision; adaptive watershed; graph cuts; energy function; stereo matching

**OCIS codes** 100.2000; 100.2960; 100.6890

## 1 引 言

近十年来, 立体匹配已经成为计算机视觉领域中的核心挑战之一。基于区域的稠密匹配算法可分为两种: 1) 局部窗口匹配法<sup>[1]</sup>, 对每个像素使用邻域(比如 7 pixel  $\times$  7 pixel), 能够使模型参数表面更灵活, 但是处理遮挡区域很困难; 2) 全局优化算法, 全局方法使用更小的邻域(常常是一对像素)去实现表

面平滑, 全局方法对遮挡区域处理得很好, 但是在平坦区域受限。

对于全局匹配方法中, 匹配算法有两类: 稀疏视差匹配和密集视差匹配。其中稀疏视差匹配<sup>[2]</sup>属于稳健性算法但是获得的结果是稀疏的, 通常是以边缘轮廓和线段等为匹配基元。密集视差匹配算法有动态规划法<sup>[3]</sup>, 信念传播(BP)<sup>[4]</sup>以及基于图割<sup>[5]</sup>等

**收稿日期**: 2012-11-07; **收到修改稿日期**: 2012-12-07

**基金项目**: 国家自然科学基金(61075011, 60675018)和教育部留学回国人员科研启动基金资助课题。

**作者简介**: 祝世平(1970—), 男, 博士, 副教授, 主要从事图像处理和视频压缩、计算机视觉和机器视觉在精密测量中的应用等方面的研究。E-mail: spzhu@163.com

多种方法。虽然这些方法会产生一些误匹配,但是几乎能够得到所有像素的视差信息。

图割是基于图论的方法,应用广泛。Szelisk等<sup>[6]</sup>把当前常用的几种能量函数最小化方法在解的质量和运行时间等方面进行了比较,发现与传统的8种算法(模拟退火算法、M-估计法等)相比,图割算法不仅总体精度高,而且在不连续区域和低纹理区域的精度也比其它算法都高。有些算法(如模拟退化算法)的精度和图割算法的精度非常接近,但后者在优化过程中收敛更快。Bleyer等<sup>[7]</sup>提出基于图像分割的立体匹配算法框架,图像分割充分利用彩色图像的颜色信息将图像分割成不同区域,有效地克服了视差图边界模糊问题,较好地处理了大的低纹理区域。

相较于已有的图割算法,本文提出的算法不需要使用明确的固定阶的约束项,这样计算量就会大大降低;同时可以保证不同类型的表面保持平滑,它既能保留目标边界信息,也能精确地保留高纹理区域。本算法还不使用传统邻域(4-邻域和8-邻域)而选用大领域来建立能量方程。大领域也能够使本算法更灵活地获得更好的目标边界。在本算法中,将相同亮度的像素当作同一个特征矢量形成像素组层。提出了一种自适应局部阈值方法,将其与传统分水岭和Prim算法相连接,来获取像素组层。在能量方程平滑项的构造上,选用大领域密集型非参数平滑模型将图像特征与深度值相联系。在能量函数最小化的处理中为了降低运算数据量,使用优化的 $\alpha$ -扩展法在像素组层范围内搜索,而不需要搜索所有像素。相较于已有的图割算法,本算法不仅对于静态图像能够获得高效率 and 高质量稠密深度图,对于动态图像即使是对每一帧图像单独处理时也能获得更加实时稳定的深度信息,动态视频帧间的矛盾也有了很大的改善,所呈现的深度视频连贯、完整,帧与帧之间不存在大片段的跳跃现象。

## 2 算法整体框架

本算法中相对于传统图割算法主要有以下改进:

1) 使用分水岭算法对图进行初始化划分,利用Prim算法将初始划分区域融合,提出一种自适应局部阈值方法,达到需求阈值停止融合,所得的最小生成树组就为像素组层;

2) 建立非参数深度平滑模型,使用大领域的密集型估计将图像特征与深度值链接,以视差组层为

标号建立能量函数;

3) 能量函数最小化使用优化的 $\alpha$ -扩展法,寻找匹配点时不需要对整个图进行搜索,而是利用像素组层区域中像素范围来搜索,在范围内搜索区域像素,寻找匹配点,否则不搜索。图1为本文算法的流程图。

## 3 算法的实现

### 3.1 自适应分水岭算法的实现

本文实现了一种优化的分水岭算法,是基于一个重复程序中自适应地产生的局部阈值,在此局部的信息也被考虑进来。这个算法包括两个主要步骤。首先,使用分水岭分割方法将图像分割成大量的区域。相邻区域图表(RAG)就是描述图像分割的数据结构。第二步是一个重复的过程,这些区域被融合且达到局部阈值停止融合。融合过程的顺序参照Prim在图表中找出最小值生长树的算法<sup>[8]</sup>,在融合过程中追踪每个区域的变化并将变化的特征保存下来。通过分析这些变化识别出在融合过程中每个区域是从何处开始变得不同类的。这样,局部阈值就产生出来了。为了取消制造不同类区域的融合,达到融合后,所有的融合从端部一个接一个的取消。这次融合涉及到的两个区域被认为是不可融合区域,即为“最后区域”。这个算法并没有任何假设参量和给定的阈值。阈值的数量和大小仅仅在程序运行结束之时才知道。通过以下方式获得的最终区域,分割结果正是图像的分区。

对于分水岭算法的输入是一个灰度的梯度图像,因此,将彩色输入图像 $I_r, I_g$ 转换为灰度图像。应用Canny边缘检测算子来得到梯度度量的图像,记为 $I_{\text{cannyleft}}, I_{\text{cannyright}}$ 。然后经过分水岭分割并投影到彩色图像上输出即是 $I_{\text{waterleft}}, I_{\text{waterright}}$ 分别分割成 $n$ 个不重叠的区域图像。当然,所得的两幅图像都属于是过分割的图像,为了得到所需要的图像,在接下来的处理中会将这些区域进行融合。以其中的 $I_{\text{waterleft}}$ 为例,将这 $n$ 个区域表示为 $N_i^q$ ,其中 $i \in [1, n]$ ,  $r_i \in [1, R_i]$ 。 $n$ 为区域数量, $R_i$ 为融合过程中 $N_i^q$ 融合的数目。在这里假设 $N_i^q$ ( $i \in [1, n]$ )是原始区域的序列,是分水岭算法在重复融合之前的输出。为了确定融合顺序,设任意两个不同区域 $N_i^q$ 和 $N_j^q$ 间相异点函数为 $f(N_i^q, N_j^q)$ 。函数 $f$ 包含色彩和边缘两个分量。对于每个区域 $N_i^q$ ,它的色调分量平均值表示为 $\mu(N_i^q)$ 。而边缘分量使用梯度大小来表示。区域 $N_i^q$ 和 $N_j^q$ 的平均梯度为 $\mu_{\text{waterleft}}(N_i^q, N_j^q)$ ,

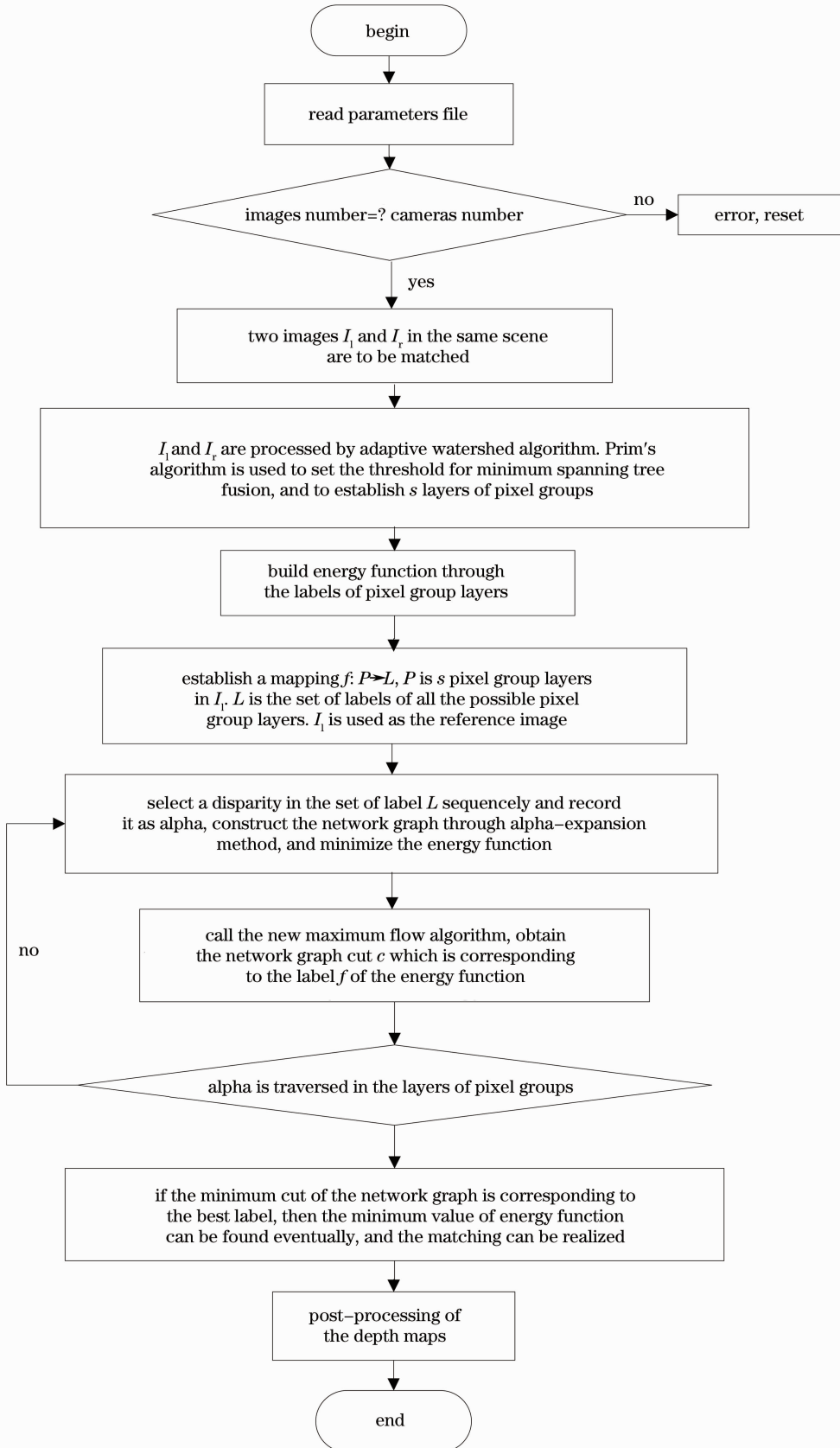


图 1 本文算法的流程图

Fig. 1 Flow chart of the proposed algorithm

是由两个区域共有的像素梯度计算出来的。图像  $I_{\text{waterleft}}$  中的像素值即为梯度值,  $M_{ij}$  表示区域  $N_i^r$  和  $N_j^r$  边界像素的集合,  $\mu_{\text{waterleft}}(N_i^r, N_j^r)$  为

$$\mu_{\text{waterleft}}(N_i^r, N_j^r) = \frac{\sum_{(x,y) \in M_{ij}} I_{\text{waterleft}}(x,y)}{|M_{ij}|}, \quad (1)$$

式中  $|M_{ij}|$  为  $M_{ij}$  中的像素数量。

$$f(N_i^r, N_j^r) = \omega_1 d[\mu_h(N_i^r), \mu_h(N_j^r)] + \omega_2 \mu_{\text{waterleft}}(N_i^r, N_j^r), \quad (2)$$

式中  $d[\mu_h(N_i^r), \mu_h(N_j^r)]$  是  $N_i^r$  和  $N_j^r$  的平均值之差, 即

$$d[\mu_h(N_i^r), \mu_h(N_j^r)] = \min\{|\mu_h(N_i^r) - \mu_h(N_j^r)|, [360 - |\mu_h(N_i^r) - \mu_h(N_j^r)|]\}. \quad (3)$$

$\omega_1$  和  $\omega_2$  是预定义的恒定系数, 相对于梯度大小来说, 非相关点函数更依赖于色调彩色空间, 所以有  $\omega_1 \gg \omega_2$ 。由于在实验中处理不同的图像, 因此在此设  $\omega_1 = 0.9, \omega_2 = 0.1$ 。

经典最小生成树方法有 Kruskal 算法和 Prim 算法<sup>[8]</sup>。通过比较, 两者都属于贪心算法, 而运行时间上, Prim 算法优于 Kruskal 算法, 虽然 Prim 算法效率没有 Kruskal 算法高, 但是 Prim 算法每次都是选择边的权值最小的节点, 而边的权值可以定义为和位置、视差以及颜色相关的量, 这样选取最小值得到的生成树符合本文算法的需要, 而且它易于实现, 所以选择使用 Prim 算法来生成最小生成树。令  $G = (V, E)$  为 RAG 结构, 表示对图像  $I_{\text{waterleft}}$  的初始

分割, 其中  $e(i, j)$  的权值为函数  $f(N_i^r, N_j^r)$  函数的值。G 的生长树为一个贯穿它所有节点的非周期连接子图。其中当  $|V| = n$  时, G 的每个生成树有  $n - 1$  个边。每个边都有 1 个权值, G 的最小生成树就是总权值之和最小的生成树。设边的集合为  $A$ 。

一般的融合方法都是只设置单个阈值, 达到这个值就停止融合, 但是这样容易造成不必要的误差, 本算法需要的区域通过单个阈值并不能得到满意的效果并会产生过多的误差, 所以提出一种自动计算局部阈值(即自适应局部阈值)的方法, 这样经过 Prim 算法后得到的每个区域由于阈值不同所以大小是不同的, 并且可以根据所需精度进行调整。由于融合不同区域时, 区域的同一性质就会产生比较大的变化, 利用这个变化来确定局部阈值, 简单来说就是融合过程中, 一旦融合的区域不同就停止融合。

令  $\mu_V(N_i^r)$  为区域  $N_i^r$  的亮度分量平均值, 且令  $V(x, y)$  为坐标  $(x, y)$  处的亮度分量值。对于任何一个区域  $N_i^r, i = 1, 2, 3, \dots, n$ , 在它的第  $r_i$  次融合时产生的变化定义为

$$\sigma(N_i^r) = \frac{1}{|N_i^r|} \sqrt{\sum_{(x,y) \in N_i^r} [V(x,y) - \mu_V(N_i^r)]^2}, \quad (4)$$

式中  $|N_i^r|$  为  $N_i^r$  的大小。

区域  $N_i^r$  在第  $r_i$  次融合时的变化为  $\Delta\sigma(N_i^r) = |\sigma(N_i^r) - \sigma(N_i^{r-1})|$ 。令  $J$  表示  $\Delta\sigma(N_i^r), r_i = 1, \dots, R_i$  的局部最大值集合, 局部阈值由下式决定:

$$\begin{cases} J_i = \{[r_i, \Delta\sigma(N_i^r)] \mid \Delta\sigma(N_i^r) > \sigma(N_i^{r-1}) \ \& \ \Delta\sigma(N_i^r) > \sigma(N_i^{r+1})\} \\ \Delta\sigma(N_i^r) > \delta, \quad \delta = \frac{1}{K_i} \sum_{r_i \in J_i} \Delta\sigma(N_i^r) \end{cases}, \quad (5)$$

式中  $\delta = \frac{1}{K_i} \sum_{r_i \in J_i} \Delta\sigma(N_i^r)$  为  $\Delta\sigma(N_i^r)$  的平均值。

产生 MST 的过程即融合区域的过程, 记述如下, 设  $T_E$  是 G 上最小生成树中边的集合:

- 1) 从  $U = \{u_0\} (u_0 \in V), T_E = \emptyset$  开始;
- 2) 在所有  $u \in U, v \in V - U$  的边  $(u, v) \in E$  中找到一条权值的边  $(u_0, v_0)$  归入  $T_E$  中, 同时  $v_0$  并入  $U$  中;
- 3) 重复执行步骤 2, 达到(5) 式所设置的阈值, 停止融合形成一棵最小生成树  $T_i$ ;
- 4) 重复步骤 2、3, 直到  $U = V$  为止。

通过上述方法使用 Prim 方法的最大生成树来

估计最初密集图, 如图 2 所示。从原始图到分为多个 T 树, 合并所有这些树去合成稀疏图以估计原始密集图, 这样一个稀疏图  $G^s$  至多有  $T(L-1)$  个边界。

### 3.2 建立能量方程

图割是用来解决优化能量函数问题, 所以必须把需要解决的问题构造成能量函数, 在所有割中找到一个具有最小权值的割, 即寻求能量函数的最小解。给出一组图像对  $I_1$  和  $I_r$ , 通过最小化下列方程  $E$  分别估算它们的视差图  $D_1, D_r$ :

$$E(D) = E_{\text{data}}(D_1, D_r) + E_{\text{smooth}}(D_1) + E_{\text{smooth}}(D_r). \quad (6)$$

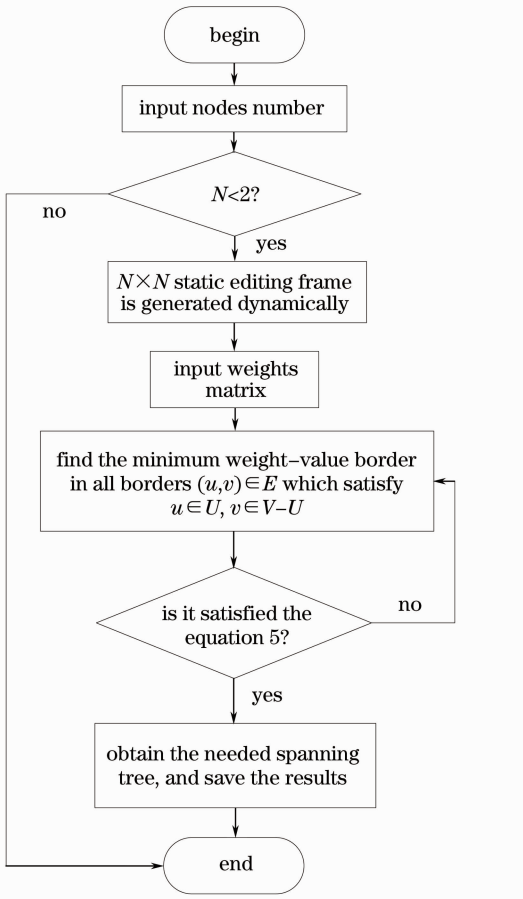


图 2 Prim 算法流程图

Fig. 2 Flow chart of Prim's algorithm

### 3.2.1 图的一致性

图的一致性有很多种方法<sup>[9]</sup>, 在这里使用文献[10]中提出的方法:

$$E_{\text{data}}(D_l, D_r) = \sum_{p \in I_l} e_{\text{data}}(d_p, d_q), \quad (7)$$

式中  $d_p = D_l(p)$  是图  $I_l$  中像素  $p$  的视差,  $q = p + D_l$  是图  $I_r$  中  $p$  的相对应的像素,  $d_q = D_r(q)$  是  $I_r$  中  $q$  的视差。 $e_{\text{data}}$  定义为

$$e_{\text{data}}(d_p, d_q) = \begin{cases} 0 & \text{if } d_p < d_q \\ \rho_{\text{data}}(\|c_p - c_q\|^2) & \text{if } d_p = d_q, \\ 0 & \text{if } d_p > d_q \end{cases} \quad (8)$$

式中  $c_p = I_l(p)$ ,  $c_q = I_r(q)$ ,  $\rho_{\text{data}}$  是图一致性的一个常数。令  $\rho_{\text{data}}(x) = \min(0, |x| - \tau_{\text{data}})$ , 对于多目立体视频, 图的一致性是在一组被选择的图像对中。

### 3.2.2 平滑项

平滑项如下:

$$E_{\text{smooth}}(D) = \sum_{p \in I} E_{\text{smooth}}(d_p; \{d_q\}_{q \in N_p}), \quad (9)$$

式中  $E_{\text{smooth}}$  模型连接像素  $p$  的视差  $d_p$  和像素  $q$  在  $p$

的邻域  $N_p$  的视差  $d_q$ 。

传统上, 邻域是在图像栅格的基础上定义的, 如 4-邻域和 8-邻域。这样定义的邻域与人眼视觉感知是不相容的, 因此邻域不应该被限制在图像栅格中。将相同亮度的像素当作同一个特征矢量形成像素组层, 用在每个像素组层的区域里的像素来定义邻域。使用 5D 矢量  $f = [x, c]$  作为特征矢量, 包括位置  $x = [x, y]$  和颜色  $c = [R, G, B]$ 。为明确叙述特征矢量的空间, 定义像素  $p$  以及其邻域  $N_p$ ,  $N_p$  包括图像中除了  $p$  以外的所有剩余像素。在邻域  $N_p$  中每个像素是一个结点并在  $p$  和  $q$  之间组成了一个权值边界  $w_{p,q}$ 。通过用一个权值  $w_{p,q} + w_{q,p}$  来连接边界  $p-q$  和  $q-p$ 。而这个权值的构建能够很好地解决匹配中的遮挡问题。寻求基于邻域  $N_p$  使用非参数归正<sup>[11]</sup> 预测视差  $d_q$ 。建立连接在  $N_p$  视差  $d$  和特征  $f$  的分类:

$$P(d, f | N_p) = \frac{1}{|N_p|} \sum_{q \in N_p} g_d \left( \frac{d - d_q}{\sigma_d} \right) \times g_x \left( \frac{x - x_q}{\sigma_x} \right) g_c \left( \frac{c - c_q}{\sigma_c} \right), \quad (10)$$

式中  $g_d$ ,  $g_x$  和  $g_c$  分别是视差  $d$ , 像素位置  $x$  和像素颜色  $c$  的核心方程。在(10)式中对像素位置  $x$  和颜色  $c$  使用高斯核心方程。 $\sigma_d$ ,  $\sigma_x$  和  $\sigma_c$  分别是相关的频带宽度。为预测  $d_p$ , 根据给定的  $f_p$  计算出  $f_p$  所有可能值<sup>[11]</sup>:

$$P(d | f_p, N_p) = \sum_{q \in N_p} w_{p,q} g_d \left( \frac{d - d_q}{\sigma_d} \right), \quad (11)$$

式中

$$w_{p,q} = \frac{g_x \left( \frac{x - x_q}{\sigma_x} \right) g_c \left( \frac{c - c_q}{\sigma_c} \right)}{\sum_{q \in N_p} g_x \left( \frac{x - x_q}{\sigma_x} \right) g_c \left( \frac{c - c_q}{\sigma_c} \right)}, \quad (12)$$

(11)式遵循了混合分配, 使用它构造  $d_q$  的归正, (9)式中  $E_{\text{smooth}}$  就变为

$$E_{\text{smooth}}(d_q; N_p) = -\lambda \lg[P(d_p | f_p, N_p)], \quad (13)$$

式中  $\lambda$  是归一化系数。通过实验, (10)式中不需要使用整幅图像去评估  $P(d_p | f_p, N_p)$ 。

注意到(13)式取决于未知变量(深度)和观测数据(像素颜色和位置)。这个能量方程在条件随机场(CRF)结构<sup>[12]</sup>中明确阐述了。(13)式能帮助找到参数  $\sigma_d, \lambda$  等。 $\lg(\cdot)$  是一个凸函数, 因此使用(11)式中的  $P(d_p | f_p)$  的表达式, 在(11)式中有以下  $E_{\text{smooth}}$  的不等式:

$$E_{\text{smooth}}(d_p; N_p) < -\lambda \sum_{q \in N_p} w_{p,q} \lg \left[ g_d \left( \frac{d - d_q}{\sigma_d} \right) \right]. \quad (14)$$

让  $\rho_{sn}(d-d_q) \stackrel{\text{def}}{=} -\lg\left[g_d\left(\frac{d-d_g}{\sigma_d}\right)\right]$ 。以(12)

式代替原始规划(11)式,得到  $E_{\text{smooth}}$  上限值  $E_{\text{smooth}}^u$ , 它只包含两个视差变量。使用  $E_{\text{smooth}}$  上限值  $E_{\text{smooth}}^u$  来代替总能量方程中的立体模型,对  $E$  有上限值  $E^u$ :

$$E^u(D_l, D_r) = E_{\text{smooth}}(D_l, D_r) + E_{\text{smooth}}^u(D_l) + E_{\text{smooth}}^u(D_r). \quad (15)$$

原则上,图割能够最优化  $E^u$ ; 因为  $E_{\text{smooth}}^u$  包括大量的对偶项,所以执行起来计算量过大,其中很多  $\omega_{p,q}$  的权值很小(比如那些包括两个位置距离很远的像素值或两个颜色值不相似的像素),所以本文中忽略这些对匹配只有微小影响的项,并且利用像素组层寻找一组优势项来接近  $E_{\text{smooth}}^u$ 。

利用 3.1 节中自适应分水岭算法所得到的像素组层来寻找这样的一组值,在像素组层组成的稀疏图中权值边界  $\omega_{p,q}$  来估计  $E_{\text{smooth}}^u$ :

$$E_{\text{smooth}}^u \approx E_{\text{smooth}}^{\text{au}}(D) = \lambda \sum_{(p,q) \in G^s} \omega_{p,q} \rho_{sn}(d_p - d_q), \quad (16)$$

式中上标“au”代表“估计的上限值”。用  $E_{\text{smooth}}^{\text{au}}$  来取代(15)式中的  $E_{\text{smooth}}^u$ , 对于(6)式中原始立体模型中的  $\Phi$  有估计上限值  $E^{\text{au}}$ :

$$E^{\text{au}}(D_l, D_r) = E_{\text{smooth}}^{\text{au}}(D_l, D_r) + E_{\text{smooth}}^{\text{au}}(D_l) + E_{\text{smooth}}^{\text{au}}(D_r). \quad (17)$$

通过多次实验,发现  $T=2$  效果很好,在  $T=2$  时,在稀疏图中有  $2L$  个边界。在利用图像栅格的 4 领域系统中有差不多同样数目的图边界。此外,(17)式通过图割方法满足公共约束。因此,使用图割可以更有效的优化(17)式。

### 3.3 匹配过程(能量函数最小化)

基于图割有两种能量最小化算法<sup>[7]</sup>:  $\alpha$ -扩展算法和  $\alpha$ - $\beta$  交换算法。这两个算法在特定情况下(如标号取 0 和 1 的时候)可以找到全局最优解。两种算法的结构是相似的,在每一次循环中,算法都以一定的顺序对每个标号( $\alpha$ -扩展算法)或每对标号( $\alpha$ - $\beta$  交换算法)进行一次迭代。

在  $\alpha$ -扩展算法的一次循环中共执行了  $|L|$  次迭代,在  $\alpha$ - $\beta$  交换算法的一次循环中执行了  $|L|^2$  次迭代,所以选择使用  $\alpha$ -扩展算法来进行优化。

匹配过程就是能量函数最小化的过程。对传统的  $\alpha$ -扩展算法进行优化。传统的  $\alpha$ -扩展算法寻找匹配点时需要对整个图进行搜索,而优化后的  $\alpha$ -扩

展算法是利用像素图层区域中像素范围来搜索,寻找匹配点,如果匹配点不在该区域内,不进行搜索,这样就大大降低了运行时间。

完成能量函数最小化的过程之后,根据得到的视差分布生成相应的深度图。

## 4 实验结果与讨论

实验平台为: Intel<sup>®</sup> Core<sup>™</sup> 2 Duo Processor CPU, 主频 1.86 GHz, 3.25 GB DDR2 内存。立体匹配系统的编程环境为 Microsoft Visual Studio 2008。使用 Middlebury 平台<sup>[13]</sup>提供的四组图像以及双目视频序列 Plant。为了验证本文算法的有效性和准确性,对多个格式的视频测试序列进行了实验。图 3 为针对 Middlebury 平台提供的四组图像进行测试,并给出了本文算法所获得的深度图。图 4 和图 5 分别为针对双目视频序列 Plant 中的第 3 帧和第 7 帧图像进行测试,采用本文算法所得到的深度图。

在图 3 中,(10)式中对像素位置  $x$  和颜色  $c$  使用高斯核心方程:  $g_x(x) = g_c(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)$ 。在  $\sigma_x \approx 20$  和  $\sigma_c \approx 5$  时对所有图像效果都是最好。

在图 3 中,4 组图像背景复杂度、高曲率表面和不连续性区域都有很大的区别,但是使用本算法所得到的深度图效果都很好,对于高曲率表面的处理很平滑,所得不连续区域的边界很清晰。

此外,标准多视图图割匹配方法<sup>[10]</sup>依赖于通过矩形图像栅格的平滑约束,这样不能将物体边界信息维持得很好,并且边界表现出过时颤抖。而其他方法<sup>[14,15]</sup>是基于图像分割来更好地保持目标的边界,但是可以看出深度图在帧与帧之间的很大区域中呈现跳跃现象。本文算法对于动态视频帧间的矛盾有了很大的改善,所呈现的深度视频连贯、完整,帧与帧之间不存在大片段的跳跃现象。

基于分割的立体匹配方法能够很好的实现平坦区域的重建。现有方法的一个潜在问题就是当应用在动态视频序列中时,图像分割在视频帧间是相互矛盾的,深度估计的结果常常呈现不连续的跳跃现象。为突出本算法所获取的深度图在帧间不呈现跳跃现象,在图 4 和图 5 中选取了双目视频序列 Plant 中第 3 帧和第 7 帧的深度图来表现。

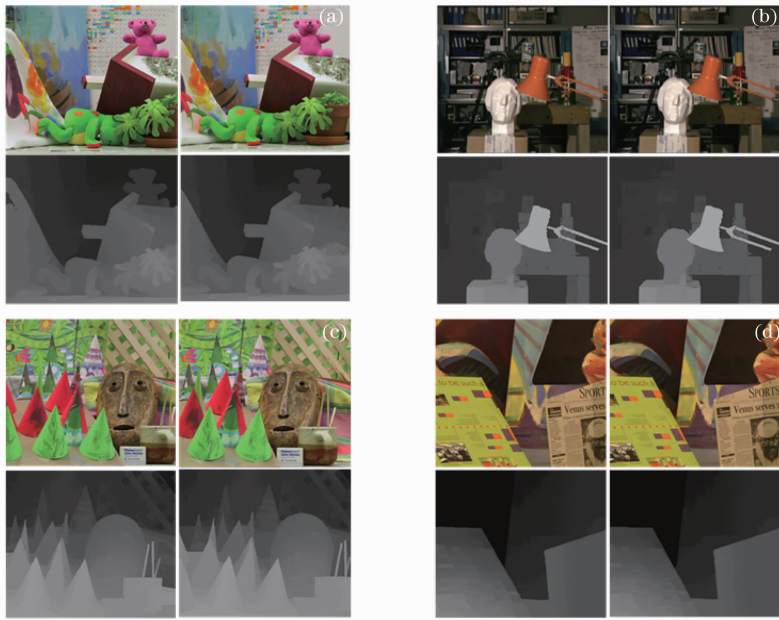


图 3 Middlebury 测试平台提供的四组测试图像及采用本文算法得到的深度图。(a) Tsukuba 左右目及深度图；  
(b) Teddy 左右目及深度图；(c) Venus 左右目及深度图；(d) Cones 左右目及深度图

Fig. 3 Stereo images provided by Middlebury test platform and the depth maps acquired by the proposed algorithm. (a) Tsukuba's left and right views and depth maps; (b) Teddy's left and right views and depth maps; (c) Venus's left and right views and depth maps; (d) Cones's left and right views and depth maps

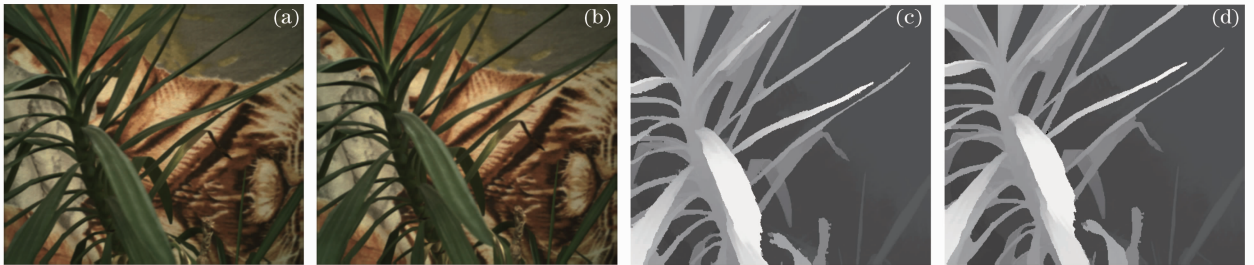


图 4 Plant 双目视频系列第 3 帧左目和右目图像及采用本文算法所得到的深度图。(a) Plant 第 3 帧左目图像；  
(b) Plant 第 3 帧右目图像；(c) Plant 第 3 帧左目深度图；(d) Plant 第 3 帧右目深度图

Fig. 4 Stereo images of 3<sup>rd</sup> frame of plant and the depth maps acquired by the proposed algorithm. (a) Left view of 3<sup>rd</sup> frame of plant; (b) right view of 3<sup>rd</sup> frame of plant; (c) depth map of left 3<sup>rd</sup> frame of plant; (d) depth map of right 3<sup>rd</sup> frame of plant

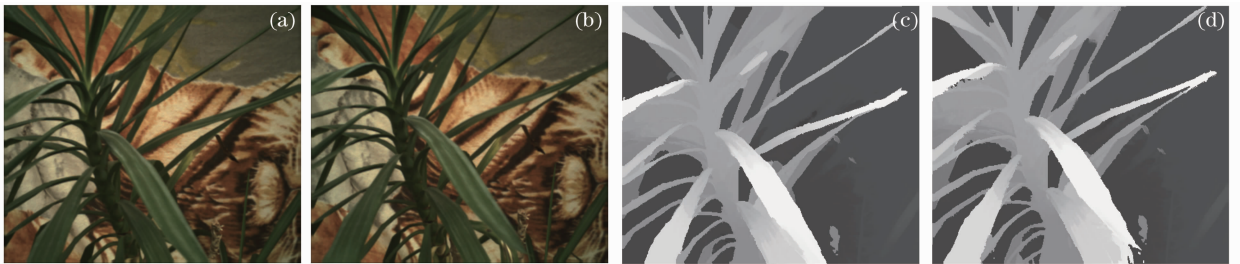


图 5 Plant 双目视频系列第 7 帧左目和右目图像及采用本文算法得到的深度图。(a) Plant 第 7 帧左目图像；  
(b) Plant 第 7 帧右目图像；(c) Plant 第 7 帧左目深度图；(d) Plant 第 7 帧右目深度图

Fig. 5 Stereo images of 7<sup>th</sup> frame of plant and the depth maps acquired by the proposed algorithm. (a) Left view of 7<sup>th</sup> frame of plant; (b) right view of 7<sup>th</sup> frame of plant; (c) depth map of left 7<sup>th</sup> frame of plant; (d) depth map of right 7<sup>th</sup> frame of plant

通过图 3、图 4 和图 5 可以看出,本算法很好的解决了现有算法中对于静态图像和动态视频深度图不能兼容的缺陷,本算法无论对于静态图像还是动态视频都有比较好的效果。稀疏图下的像素匹配可以使不连续的边界保留得很好而不需要图像分割作为预处理项。分割应用在视频中时常常出现暂时的不相容性。本方法避免了在预处理中使用分割,并且在动态视频序列中能够恢复实时稳定的深度值,在每一帧图像单独处理时也能得到很好的效果。

本算法的约束项模型是大领域无参数的。无参数模型将图像特征用深度值表示时不需要使用明确的固定阶的约束项,这样计算量就会大大降低。大

表 1 基于 Middlebury 平台本文算法与其它算法结果的比较

Table1 Comparison results of the proposed algorithm and other algorithms based on Middlebury test platform

Algorithms	Teddy			Cones		
	All	Non-occ	Disc.	All	Non-occ	Disc
Mean split <sup>[19]</sup>	10.2	6.99	12.0	7.64	5.60	13.6
SAD <sup>[17]</sup>	27.2	19.6	34.0	22.6	13.5	22.0
AdaptBP <sup>[16]</sup>	7.06	4.22	11.8	7.92	2.48	7.32
Double BP <sup>[18]</sup>	9.24	2.90	7.82	8.71	3.55	9.70
Ours	8.02	7.78	12.2	7.84	3.05	8.21

Note: all means all regions, non-occ means the non-occluded area, disc means depth discontinuity regions. The match result has more than 2 pixels difference which means matching error.

## 5 结 论

在特征区域使用大领域非参数平滑约束模型将图像特征用深度值表示时,不需要使用明确的固定阶的约束项,使得计算量大大降低。同时可以保证不同类型的表面保持平滑,它既能保留目标边界信息,也能精确地保留高纹理区域。大邻域也能够使本文算法更灵活地获得更好的目标边界。

此外,在动态视频序列中与其它方法相比较,本算法即使对每一帧图像单独处理时也能获得更加实时稳定的深度信息。本算法对于动态视频帧间的矛盾也有了很大的改善,所呈现的深度视频连贯、完整,帧与帧之间不存在大片段的跳跃现象。

通过 Middlebury 平台对算法定量评估得出在所有区域的误匹配率、非遮挡区域以及深度不连续区域的误匹配率都控制在 8.5% 以内,在 Middlebury 平台 135 组排行数据中排名第 19 位。

## 参 考 文 献

- 1 Song Zhendong, Jiang Wei. Monocular multi-viewpoints stereo imaging and application [J]. *Acta Optica Sinica*, 2012, **32**(5): 0515001
- 宋振东,姜 伟. 单目多视点立体图像提取及应用[J]. *光学学*

报, 2012, **32**(5): 0515001

邻域能够使本文算法更灵活地获得更好的目标边界。本算法能够在不连续的边界和高折叠纹理区域都得到很好的效果。

由表 1 可以得出,本算法属于全局优化算法,与局部优化算法均值分割<sup>[16]</sup>和绝对值和(SAD)<sup>[17]</sup>算法相比较,各项错误率均明显低于局部优化算法。而在同类算法中,本算法的错误百分比略高于 Middlebury 平台上的最优算法 AdaptBP<sup>[16]</sup>(排名第 4),但部分错误百分比低于 Middlebury 平台上的最优算法 Double BP<sup>[18]</sup>(排名第 6),本算法在对 Middlebury 平台提供的四组图像处理,实验结果在 Middlebury 平台 135 组排行数据中排名第 19 位。

- 2 J. Cech, R. Sara. Efficient sampling of disparity space for fast and accurate matching [C]. *IEEE Proceedings of the Second International ISPRS Workshop (BenCOS 2007)*, Minneapolis, MN, USA, June 23, 2007, 1~8
- 3 Liu Yamei, Chang Chengguo. The practical application of dynamic programming and network flow algorithm [J]. *Software Guide*, 2012, **11**(7): 20~23
- 刘雅梅,常呈果. 动态规划和网络流算法的实际应用[J]. *软件导刊*, 2012, **11**(7): 20~23
- 4 Zhou Ziwei, Fan Jizhuang, Zhao Minjie *et al.*. Parallel stereo matching algorithm base on belief propagation [J]. *Optics and Precision Engineering*, 2011, **19**(11): 2774~2781
- 周自维,樊继壮,赵晋杰等. 基于置信传播的立体匹配并行算法[J]. *光学精密工程*, 2011, **19**(11): 2774~2781
- 5 P. F. Felzenszwalb, R. Zabih. Dynamic programming and graph cut algorithms in computer vision [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(4): 721~740
- 6 R. Szelisk, I. R. Zabih. An experimental comparison of stereo algorithms [C]. *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, 2000. 1883: 1~19
- 7 M. Bleyer, M. Gelautz. Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions [J]. *Signal Processing: Image Communication*, 2007, **22**(2): 127~143
- 8 T. H. Cormen, C. E. Leiserson, R. L. Rivest *et al.*. *Introduction to Algorithms* (2nd Ed.) [M]. MIT Press, 2001. 348~352
- 9 D. Scharstein, R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. *International*



- J. *Computer Vision*, 2002, **47**(1-3): 7~42
- 10 V. Kolmogorov, R. Zabih. Multi-camera scene reconstruction via graph cuts [C]. *Proceedings of the Seventh European Conference on Computer Vision (ECCV 2002)*, 2002, 3: 82~96
- 11 C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*[M]. Springer, 2006
- 12 J. D. Lafferty, A. McCallum, F. C. N. Pereira. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]. *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)*, 2001. 282~289
- 13 Daniel Scharstein, Richard Szeliski. Middlebury stereo vision [OL]. <http://vision.middlebury.edu/stereo/>
- 14 O. J. Woodford, P. H. S. Torr, I. D. Reid *et al.*. Global stereo reconstruction under second order smoothness priors [C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, 2008. 2115~2128
- 15 L. C. Zitnick, S. B. Kang, M. Uyttendaele *et al.*. High-quality video view interpolation using a layered representation [J]. *ACM Transactions on Graphics*, 2004, **23**(3): 600~608
- 16 A. Klaus, M. Sormann, K. Karner. Segment based stereo matching using belief propagation and a self adapting dissimilarity measure [C]. *Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006)*, 2006. 15~18
- 17 Sang Hwa Lee, Siddharth Sharma. Real-time disparity estimation algorithm for stereo camera systems [J]. *IEEE Transactions on Consumer Electronics*, 2011, **57**(3): 1018~1026
- 18 Q. Yang, R. Yang, J. Davis *et al.*. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling [J]. *Pattern Analysis and Machine Intelligence*, 2009, **31**(3): 492~504
- 19 Wu Chunhong, Fu Guoliang. A stereo matching method based on K-means segmentation and neighborhood constraints relaxation [J]. *Chinese J. Computers*, 2011, **34**(4): 755~760
- 伍春洪, 付国亮. 一种基于图像分割及邻域限制与放松的立体匹配配方[J]. *计算机学报*, 2011, **34**(4): 755~760

栏目编辑: 李文喆