基于时空测地线传播的 RGB-D 视频分割

王 斌[1)], 陈文拯[1)], 钟 凡[1)*], 屠长河[1)], 秦学英[1)], 彭群生[2)]

[1)] (山东大学计算机科学与技术学院 济南 250101)
[2)] (浙江大学 CAD&CG 国家重点实验室 杭州 310027)
(zhongfan@sdu.edu.cn)

摘 要: 针对前景和背景深度交叠或相机运动时基于深度统计的传统视频分割算法中存在的问题, 提出一种基于时空测地线的方法, 并证明该方法适合基于深度的视频分割. 首先使用基于运动检测的方式进行初始化; 然后使用基于特征点选择方式定义种子结点, 特征点匹配方式构建时域链接, 空间上 8 邻域像素连接形成空域链接, 在连续两帧之间构建时空测地线传播图; 最后在时空测地线传播图上使用泛化测地线距离变换将前一帧的分割结果传播到当前帧, 并自适应地在传播和检测间切换消除累计误差. 实验结果表明, 该方法能够在复杂场景和相机运动情形下输出稳定的分割结果.

关键词: 视频分割; 深度图; 测地距离; 分割传播
中图法分类号: TP391.41

RGB-D Video Segmentation via Geodesic Spatio-Temporal Propagation

Wang Bin[1)], Chen Wenzheng[1)], Zhong Fan[1)*], Tu Changhe[1)], Qin Xueying[1)], and Peng Qunsheng[2)]

[1)] (*School of Computer Science and Technology, Shandong University, Ji'nan* 250101)
[2)] (*State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou* 310027)

Abstract: Traditional video segmentation methods based on depth statistics often fail in case of moving camera or overlap between depth ranges of foreground and background. In this paper we propose a geodesic-based method that is suitable for RGB-D video segmentation. The segmentation process is initialized by motion detection, then for each two consecutive frames, a geodesic spatio-temporal graph is constructed, with seed nodes selected based on image feature detection, temporal links built via feature matching and the 8 spatial neighborhoods of pixels used as spatial links. The segmentation result of previous frame then is propagated to the current frame via geodesic spatio-temporal propagation, which is conducted efficiently by generalized geodesic distance transform. Accumulated errors are eliminated by alternatively launching the process of propagation and motion detection. Experiments demonstrate the robust segmentation results for videos of complex scenes and moving camera.

Key words: video segmentation; depth map; geodesic distance; segmentation propagation

视频分割是基础的具有挑战性的课题. 视频中动态对象的分割被应用到诸多实际工程, 如视频监控、电影制作和增强现实应用[1]. 在没有高层知识指导的情况下, 视频分割本质上是困难并有

收稿日期:2014-09-10;修回日期:2015-03-17. 基金项目:国家自然科学基金(61332015, 61173070, 61202149, 61303089); 山东省优秀中青年科学家奖励基金(BS2013DX011). 王 斌(1990—), 男, 博士研究生, 主要研究方向为目标识别与检测; 陈文拯(1992—), 男, 硕士研究生, 主要研究方向为图像视频处理; 钟 凡(1982—), 男, 博士, 论文通讯作者, 主要研究方向为图像视频处理; 屠长河(1968—), 男, 博士, 教授, 博士生导师, 主要研究方向为计算机图形学; 秦学英(1966—), 女, 博士, 教授, 博士生导师, 主要研究方向为增强现实; 彭群生(1947—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究方向为真实感绘制、虚拟现实.

歧义的. 难点主要源自 2 个方面：1) 前景和背景认知或界定的歧义性，2) 前景和背景间存在不可分的统计信息.

因缺乏指导线索和计算能力的限制，实时视频分割极其困难. 现有技术或多或少地依赖静态背景假设，包括背景相减[2-3]和多特征融合技术[4-5]. 相机抖动、前景阴影、非静态背景和光照变化等情况[6] 出现时，静态背景假设极易被违背. 目前，真正稳定实用的方法还未见报道. 非静态背景和相机运动的情况下，实时视频分割会更棘手.

仅使用颜色信息的分割方法对低对比度颜色边缘和不可分颜色统计[7]非常敏感. 为此，Gordon 等[8]首次提出联合使用深度和颜色信息进行分割，深度数据通过立体相机获取. 近年来，廉价高效的 RGB-D 相机出现并得到广泛关注，例如微软的 Kinect 和华硕的 Xtion，RGB-D 相机能提供室内环境(受深度传感器限制)的实时深度和彩色数据. 因此，计算机视觉领域的一些课题，例如人体动作识别和视频分割等，逐渐开始利用 RGB-D 数据. 深度信息是一种有效指导分割的线索，它不会面临基于颜色信息的分割方法的困境，例如光照变化和不可分颜色统计.

基于深度信息的视频分割取得了一定程度的进展. Camplani 等[9]提出通过结合多个颜色和深度分类器，深度被用来增强背景相减的方法. 在没有深度传感器的情况下，Kolmogorov 等[10]利用立体匹配似然(视差统计信息)进行实时视频分割. 文献[11-12]中，深度数据源自 TOF 相机. Crabb 等[11]定义深度阈值用以提取前景，然而基于深度阈值的方法非常脆弱并且不能处理时间序列上的前景深度发生变化的情形. Wang 等[12]提出一种更自适应的方法，将前景和背景深度建模成高斯混合模型，该方法类似于对颜色分布进行建模的方法，显然，当前景和背景的深度范围发生交叠时会产生错误.

因此，即使深度信息可用，实时视频分割仍然充满挑战. 之前的方法没有有效地使用深度信息指导分割. 已有的方法或对非静态背景和相机运动敏感(基于背景相减的方法)，或不能有效处理前景和背景深度交叠的情况(基于深度统计的方法)，而现实环境中上述 2 种情况会经常出现. 本文提出一种基于时空测地线的方法，更有效地利用了深度信息，能够在复杂场景和相机运动情形下输出稳定的分割结果.

1　时空测地线传播

时空测地线传播目标是在给定前一帧分割结果后分割出当前帧，此过程被称为分割传播，交互视频分割领域的文献[7, 13]已经对它进行了深入的研究. 与基于检测方法例如背景相减相比，基于传播的方法更适合处理非静态背景和相机运动情形下的视频分割. 然而，由于前景对象内部杂乱的颜色边缘和前景的拓扑变化，加上遮挡的出现和消失，仅仅用颜色信息很难实现稳定的传播. 相比彩色图像，深度图有一种极佳的性质：一个对象内部的深度极少出现不连续的情况，深度不连续的情况仅出现在前景边界附近. 此性质使得测地线方法适合基于深度的视频分割.

1.1　时空传播图

为了将分割结果从前一帧传播到当前帧，构建时空传播图，如图 1 所示.原始的测地线图像分割方法中，每个像素与它空域上的近邻像素连接[14]. 然而，图 1 中前一帧的某些像素也被包含到时空传播图中，作为传播的种子结点. 因此，时空传播图中的边被分为：空域链接和时域链接 2 类.

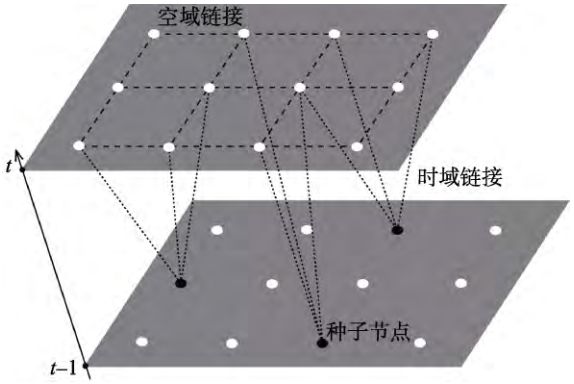


图 1　时空传播图

种子结点. 前一帧的部分像素被指定为种子结点，它们的分割状态已知，因此种子结点是分割传播的源头. 虽然理论上前一帧的所有像素都能被指定为种子结点，但是考虑到计算效率和计算稳定性，只有部分特定的像素被选作种子结点. 同时，为了确保每个局部区域被种子结点覆盖，本文采用 Shi 等[15]提出的方法为每个像素 x 计算 2 个梯度特征值 λ_1 和 λ_2 . 记较小的特征值为 $\lambda_{\min}$，Shi 等[15]指出具有较大 $\lambda_{\min}$ 的像素是判别性强的特征点. 为了选择种子结点，图像空间被均分成 10×10 的网格. 在每个网格中，具有最大 $\lambda_{\min}$ 的像素被选作

种子结点.

时域链接. 时域链接连接种子结点和当前帧的像素. 直观上, 本文采用特征跟踪的方式连接相邻帧之间匹配的像素点对. 然而, 如此构建时域链接会受到特征跟踪误差的干扰, 特别是当前景对象快速运动时. 因此, 本文为每个种子结点构建一个模糊时域链接集, 每个种子结点连接当前帧的 K 个像素, 链接的权值度量其相似性. 本文中 K 被设置为 5.

对种子结点 s, 通过计算彩色图像和深度图上匹配误差的平方和(Sum of Squared Differences, SSD), 寻找与之匹配的当前帧上的点集. 匹配误差度量方式为

$$\varepsilon(s,x)=\omega_{\mathrm{c}}\mathrm{SSD}_{\mathrm{c}}(s,x)+(1-\omega_{\mathrm{d}})\mathrm{SSD}_{\mathrm{d}}(s,x).$$

其中, x 是当前帧搜索域中的一个像素; 本文中搜索域是以 s 坐标位置为中心大小为 21×21 的窗口; $\mathrm{SSD}_{\mathrm{c}}$ 和 $\mathrm{SSD}_{\mathrm{d}}$ 分别是彩色图像和深度图的平方误差和, 本文中 SSD 的窗口大小设置为 5×5; ω_{c} 用来调节彩色数据和深度数据的权重, 当深度数据与彩色数据的取值范围一致时, ω_{c} 被设置为 0.6.

给定所有链接 (s,x) 的匹配误差 $\varepsilon(s,x)$, 它们被升序排列. 为了选择 K 个像素构建时域链接, 本文可以简单地选择前 K 个具有较小匹配误差的链接. 然而, 如此选择的链接往往聚集在同一位置附近. 因此, 任何一对链接的空间距离被规定不小于阈值 $d_{\min}$, 本文中 $d_{\min}$ 被置为 5. 通过遍历已排序的链接队列, 并丢弃不满足条件的链接, 本文得到 K 个链接作为时域链接.

测地线距离变换中, 边的权值表示 2 个结点的距离. 因此, 匹配误差小的点对对应的时域链接会被赋予较小的权值. 每条时域链接的权值定义为

$$\omega^{\mathrm{t}}(s,x)=\beta[\varepsilon(s,x)]^{p};$$

其中, β 是尺度因子; p 控制匹配误差的影响. 本文方法对 β 和 p 的取值不敏感, 其中, $\beta\in[3,10]$, $p\in[0.5,1]$, 所有的测试视频都能被有效地分割.

空域链接. 每个空域链接连接当前帧上的 2 个相邻像素, 当前帧的每个像素与它的 8 邻域像素连接. 空域链接的权值可以直观地设置为像素对的颜色和深度差. 本文发现, 在测地线分割方法中, 深度数据通常比颜色数据更稳定可靠. 与彩色图像相比, 深度图有一个显而易见的优势, 物体内部的深度往往是平滑过渡的. 相反, 由于纹理的变化或阴影影响, 彩色图像上物体内部边缘会对分割传播造成强烈干扰.

为了利用深度数据的优势, 对颜色和深度数据自适应加权是必要的, 而且, 权值应该根据空间位置的不同而变化, 而不是全局地使用统一的权值. 因此, 空域链接的权值定义为

$$\omega^{s}(x,y)=\tau_{xy}\,\|D_x-D_y\|^2+(1-\tau_{xy})\,\|I_x-I_y\|^2.$$

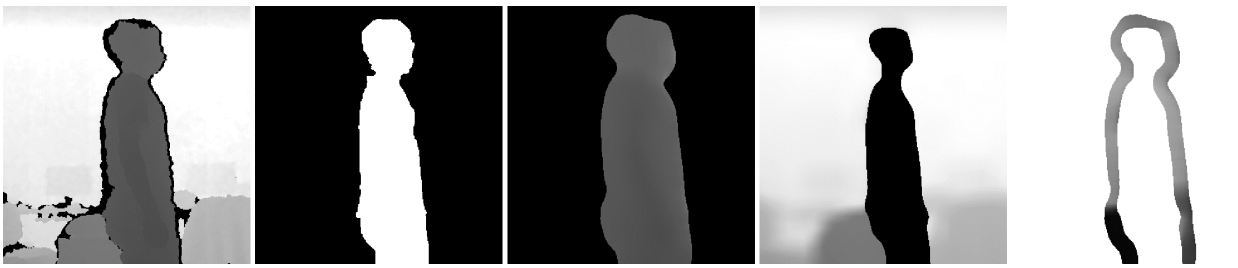
其中, D 和 I 分别是深度图和彩色图像; 权值函数 τ_{xy} 采用自适应加权策略定义, 当深度数据可用时深度被赋予较高的权值, 只有当深度数据缺失或在物体边界附近前景和背景深度接近时, 颜色被用作加权的补偿信息. 前一帧的分割结果被用来定义权值函数

$$\tau_{xy}=1-\mathrm{e}^{\frac{-\min(C_x,C_y)^2}{\sigma_{\tau}^2}};$$

其中 C 是深度置信度, 定义为

$$C_x=[x\notin\phi]\left(\frac{1}{\gamma}|\bar{D}_x^{\mathrm{F}}-\bar{D}_x^{\mathrm{B}}|\right)^{2[x\in\theta]} \tag{1}$$

ϕ 表示缺失深度数据的像素集; $[\cdot]\in\{0,1\}$ 是指示函数. 深度传感器 Kinect 输出的深度图包含部分深度值缺失的区域, 这些区域的深度置信度将被设为 0. 式(1)第 2 项度量了同一位置上前景深度和背景深度的差异, 如图 2e 所示, 差异越小, 深度置信度越小, 如此便于处理前景和背景的某些部分有相同或相近深度的情形. 用 $W\times W$(本文 W 取值为



a. 输入深度图 b. 前景掩码图 c. 前景深度均值图 d. 背景深度均值图 e. 边界深度差异图

图 2 深度置信度说明图

31)的均值滤波器处理前一帧的前景区域得到膨胀了的前景深度均值图 $\bar{D}^{\mathrm{F}}$, 如图 2c 所示, $\bar{D}^{\mathrm{F}}$ 中前景区域被膨胀了约 W 像素. 用同样均值滤波器处理前一帧中的背景区域得到背景深度均值图 $\bar{D}^{\mathrm{B}}$,如图 2d 所示. $\gamma=200$ 是正则化因子, θ 表示 $\bar{D}^{\mathrm{F}}$ 和 $\bar{D}^{\mathrm{B}}$ 的重叠区域, 因此若 $x\notin\theta$, 则 C_x 简化为 $[x\notin\phi]$.

1.2　泛化测地线距离变换

原始测地线分割方法[14,16]和它的泛化形式[17]在 2 维图像空间定义. 上述方法可以很直接地被扩展到 3 维图像序列空间, 本文面临的问题并不需要如此操作. 图 1 定义的传播问题可以高效地使用 2 维泛化测地线距离变换解决.

为了完整地描述方法流程, 本文简述泛化测地线距离变换. 给定一个定义在 2 维图像空间的距离函数 ∇_{ij}, i 和 j 是一对邻域像素, ∇_{ij} 表示它们之间的距离. E 是一个实数场, 为每个像素赋予一个软掩码值 E_x. 泛化测地线距离场可以定义为

$$G(x;\nabla,E)=\min_{x'}(d(x,x')+\nu E(x'));$$

其中, $d(x,x')=\min\limits_{[x,x']}\sum\limits_{(i,j)\in[x,x']}\nabla_{ij}$, $[x,x']$ 是任意一条连接 x 和 x' 的路径, (i,j) 是路径上的一条边. 上述问题可以通过快速 2 维测地线距离变换[18]求解.

第 1.1 节定义的传播问题使用泛化测地线距离变换求解, 空域链接的权值作为 ∇_{ij} (即 $\nabla_{ij}=\omega^s(i,j)$), 时域链接的权值被变换成 E,

$$E^l(x)=\min_{s\in S^l}\{\{\omega^{\mathrm{t}}(s,x)\},\mathrm{INF}\},l\in\{\mathrm{F},\mathrm{B}\};$$

其中, S^l 是前景或背景种子结点集, INF 表示无穷大实数, l 是前景标签 F 或背景标签 B. 第 1.1 节中, x 或与多个种子结点连接, 或与种子结点没有连接. 前一种情况下, $E^l(x)$ 取值为与 x 连接的所有时域链接的权值的最小值; 后一种情况下, $E^l(x)$ 取值 INF. 求解得到前景距离场 $G(x;\nabla,E^{\mathrm{F}})$ 和背景距离场 $G(x;\nabla,E^{\mathrm{B}})$ 后, 通过比较每个像素的前景测地线距离和背景测地线距离, 能初步获得分割结果. 当 $G(x;\nabla,E^{\mathrm{F}})<G(x;\nabla,E^{\mathrm{B}})$ 时, 像素 x 标记为前景; 否则, 被标记为背景.

按上述方法可以得到二分割结果. 然而, Kinect 提供的深度数据包含噪声, 在物体边界附近的二分割结果会出现小瑕疵, 分割结果有轻微闪烁. 完全去除上述瑕疵非常困难, 尤其当前景快速运动时, 彩色图像和深度图之间不能被精确对齐. 因此, 前景边界附近的分割结果被简单地平滑以减弱闪烁现象.

2　基于时空测地线传播的视频分割

时空测地线传播将前一帧的分割结果传播到当前帧, 因此用来分割视频时, 需要一种简易的方式初始化最初的分割状态. 为了获得精确稳定的分割结果, 累积误差去除和边界附近的分割结果细化也是必需的.

2.1　基于运动检测的自动初始化

简单可靠的初始化对视频分割是至关重要的. 基于颜色的视频分割方法通常采用背景相减或提示用户提供第一帧的分割结果的方式初始化. 在深度数据的辅助下, 本文采用基于运动检测的自动初始化方法.

在初始化阶段相机保持静止, 同时前景保持简单的运动状态. 为了提取前景对象, 本文首先基于相邻帧彩色数据和深度数据的差异执行运动检测, 然后利用运动检测的结果提取特征点. 记 $\hat{S}$ 为提取的特征点集,

$$M_x=\|I_x^t-I_x^{t-1}\|^2+\|D_x^t-D_x^{t-1}\|^2,x\in\hat{S}$$

是特征点 x 的运动量. 给定背景运动量的阈值 T_0, 背景软掩码图 $\hat{E}^{\mathrm{B}}$ 定义为

$$\hat{E}_x^{\mathrm{B}}=\begin{cases}\lambda\mathrm{e}^{\frac{-(T_0-M_x)^2}{\sigma_0^2}}, & x\in\hat{S}\text{ and }M_x<T_0;\\ \mathrm{INF}, & \text{otherwise}\end{cases}$$

前景软掩码图 $\hat{E}^{\mathrm{F}}$ 类似地定义为

$$\hat{E}_x^{\mathrm{F}}=\begin{cases}\lambda\mathrm{e}^{\frac{-(M_x-T_1)^2}{\sigma_1^2}}, & x\in\hat{S}\text{ and }M_x>T_1.\\ \mathrm{INF}, & \text{otherwise}\end{cases}$$

其中, T_1 为前景运动量阈值. 本文中, $T_0=300$, $T_1=1000$, $\lambda=1000$, σ_0 和 σ_1 均被置为 20.

初始化用的 2 维距离函数 $\hat{\nabla}$ 可以通过第 1.2 节描述的方式得到. 但是在基于运动检测的初始化阶段, 式(1)的第 2 项是不能得到的, 因此本文简化式(1)为 $\hat{C}_x=[x\notin\phi]$, 以去除深度图深度缺失部分的影响. 在初始化阶段, 前景对象不能离背景对象太近, 否则部分背景会被错误地分割为前景.

给定 $\hat{E}^{\mathrm{F}}$, $\hat{E}^{\mathrm{B}}$ 和 $\hat{\nabla}$, 依第 1.2 节所述, 泛化测地线距离变换可以被用来实现分割. 然而, 基于运动检测的方法只在静态背景和运动前景的情况

有效工作，因此本文不是简单地使用视频序列的第 1 帧作为初始化帧. 为了选择合适的初始化帧，本文评价前景种子结点的质量

$$\kappa = \frac{|\{x \in \hat{S}, M_x > T_1\}|}{|\hat{S}|},$$

其中，$|\cdot|$ 表示集合的势；κ 表示运动的前景种子结点的比例. 给定 κ，若 $0.8\kappa_{\max} \leqslant \kappa \leqslant 0.9$，则此帧被选作初始化帧，其中 $\kappa_{\max}$ 是系统启动后一段时间(本文中设置为 5 s)内 κ 的最大值. κ 的值越大表明场景静止的可能性越小，小的 κ 表示场景可能处于静止状态或仅有部分物体处于运动状态，κ 小的帧不适合作分割的初始化帧.

2.2 传播与检测切换策略

基于传播的方法，分割误差会随着传播的进行而逐渐累积. 与传统的基于颜色的分割方法相比，虽然在深度数据的帮助下传播的分割误差已经被大幅减小. 但为了增加方法稳定性，通过连续监测 κ 系统会自适应地选择执行传播或检测. 当基于运动检测的初始化条件满足时，系统执行第 2.1 节的初始化方法；否则，执行时空测地线传播方法. 为了处理前景对象大小变化的情形，$\kappa_{\max}$ 会随新的分割结果的出现被更新. 经过上述处理，本文方法能处理相机运动的情形并减小累积误差的干扰. 一旦相机静止，系统执行基于运动检测的初始化，因此累积误差会被消除.

3 实验及结果分析

本文实验中的测试视频用 Kinect 拍摄，彩色图像和深度图序列由 Kinect 自动对齐，分辨率均为 640×480. 实验环境为一台配置 2.9 GHz CPU, 4 GB 内存的计算机. 图 3 展示了 4 个测试实例，第 1 行为彩色图像，第 2 行展示了彩色图像对应的深度图.

a. 深度变化 b. 深度交叠 c. 动态背景 d. 运动相机

图 3 部分测试视频

3.1 结果和对比

本文方法能处理挑战性的测试视频，对于前景深度范围变化、非静态背景和前景背景深度交叠等情形下拍摄的视频，能给出良好的分割结果.

图 4 展示了本文方法处理前景深度范围变化的视频得到的结果，并与基于深度阈值的方法[11]进行了对比. 显然，基于深度阈值的方法仅仅在前景深度在整个视频序列中基本保持不变的情形下有效工作，相比而言，本文方法更灵活.

a. 输入图像 b. 本文分割结果 c. 深度阈值的结果

图 4 深度变化的视频分割结果对比

图 5 给出了对前景背景深度交叠的视频的分割结果，视频中前景(人)的深度和背景中部分物体(家具)的深度产生了交叠. 可以看出，基于深度统计的方法[12]不能正确处理这类情形，而本文方法可以正确分割前景背景. 文献[12]方法同时使用了颜色和深度统计，图 5 展示的结果仅使用了深度统计. 为了得到合理的对比结果，在时空测地线传播时也仅使用深度数据. 对比结果清晰地证明了本文方法能更有效地利用了深度数据.

仅使用颜色信息进行视频分割时，相机运动情况下拍摄的视频很难得到很好的处理. 相机运动时，前景对象的深度变化速度快、幅度大，背景的深度和颜色的分布也发生较大程度的变化，因此准确建模深度和颜色统计信息难度变大. 时空测地线传播不会受到此类问题的影响，因为它仅用相邻帧的特征点匹配信息构建模糊时域链接集，

a. 输入图像 b. 本文分割结果 c. 深度统计的结果

图 5 深度交叠的视频分割结果对比

即使深度信息未知，仅用颜色信息进行相邻帧的特征跟踪也是可行的．图 6 展示了本文方法对运动相机拍摄的视频的处理结果．处理包含动态背景的视频本质上和处理运动相机拍摄的视频的原理一样．图 7 给出了对包含动态背景的视频的分割结果．一旦完成初始化，系统在非静态背景下传播分割结果．因此，本方法能处理动态背景和运动相机的情形．

图 6　运动相机的视频分割结果

图 7　动态背景的视频分割结果

3.2　计算性能

本文方法的视频分割速度约为 12 帧/s，表 1 给出了图 3 中 4 个测试视频的相关数据．

表 1　测试视频相关数据和视频分割速度统计

测试视频	帧数	处理时间/s	处理速度/(帧/s)
深度变化	570	55.0	10.36
深度交叠	812	63.8	12.73
动态背景	1 510	123.4	12.24
运动相机	881	78.6	11.21

3.3　算法局限性

基于时空测地线的分割方法的局限性在于它严重依赖前景对象边界附近图像的对比度．使用深度数据时，低对比度边界出现的概率被极大减少，但当彩色图像和深度图上前景对象边界的对比度都较低时，本文方法不能正确地工作，如图 8 所示．低对比度的前景对象边界需要高层特征或形状先验的指导才能被正确分割．

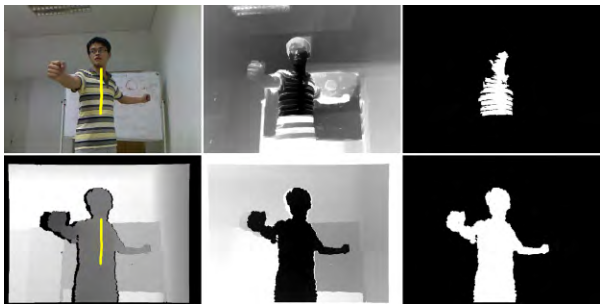
图 8　低对比度边界导致的失败结果

另一种典型的失败情形是由前景对象的快速运动导致的．前景的快速运动导致深度图和彩色图像出现较大的对齐误差，因此会产生较大的分割错误，如图 9 所示．由于 Kinect 深度传感器自身的局限，深度图和彩色图像不能完全同步获取，前景对象剧烈地移动时，深度图和彩色图像会呈现较大的对齐误差．

图 9　深度图和彩色图像没有对齐导致的失败结果

3.4　分　析

视频分割方法使用深度的方式和使用颜色的方式本质上是相同的，特别是仅用深度统计信息时．如前文所述，深度图不同于彩色图像，它不会受到物体颜色或纹理变化的干扰，所以一个物体内部深度图是平滑过渡的，对物体分割它是一个极佳的性质．本文给出一个简单的测试进行证明．本文使用基于测地线的方法分别在彩色图像和深度图上进行了粗粒度的分割．如图 10a 所示，黄色笔刷在彩色图像(第 1 行)和深度图(第 2 行)上标记了前景种子结点；图 10b 分别展示了到种子结点的测地线距离图，进行简单阈值化后得到图 10c；图 10c



a. 输入图像　　b. 测地线距离图　　c. 阈值化后分割图

图 10　彩色图像和深度图上测地线膨胀结果对比

中，深度图上的分割结果基本正确，然而彩色图像上的传播由于物体内部颜色边缘的影响而过早地终止，导致分割结果不能完整包含前景对象．本文方法利用深度图的这种平滑过渡性质，能得到优于之前方法的结果．

Kinect 提供的深度图包含噪声和深度缺失的区域，本文方法没有处理深度图上的噪声，因此分割结果受到噪声的影响．为此，可以选择更好的深度传感器，如 TOF 相机，也可以对深度图进行预处理，例如深度图修复和增强[19-20]．

4 结 论

本文提出了一种基于时空测地线传播的 RGB-D 视频分割方法．该方法的核心是时空测地线传播，它基于前一帧上稀疏的特征点分割当前帧．空域链接和时域链接同时使用深度和颜色信息构建，由于物体内部的深度通常具有平滑过渡的性质，深度被赋予了更高的权重．时空测地线传播通过泛化测地线距离变换求解．本文使用基于运动检测方式初始化时空测地线传播，同时也消除传播过程中的累积误差．实验结果表明，本文方法灵活稳定，能处理相机运动和前景背景深度交叠等情形．最后讨论分析了该方法的局限性，给出了失败的测试结果．未来的工作是关注噪声深度数据下高质量的实时视频分割，同时考虑更加合理使用 RGB-D 数据．随着新的深度传感设备不断涌现，如 Kinect2，基于 RGB-D 数据的视频分割会得到更多的关注．

参考文献(References)：

[1] Zhang Yijiang, Qin Xueying, Julien Pettré, *et al*. Inserting virtual characters into live video of real scenes[J]. Journal of Computer-Aided Design & Computer Graphics, 2011, 23(1): 185-191 (in Chinese)
(张艺江, 秦学英, Julien Pettré, 等. 虚拟群体与动态视频场景的在线实时融合[J]. 计算机辅助设计与图形学学报, 2011, 23(1): 185-191)

[2] Han B, Davis L S. Density-based multifeature background subtraction with support vector machine[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(5): 1017-1023

[3] Cheng L, Gong M L, Schuurmansc D, *et al*. Real-time discriminative background subtraction[J]. IEEE Transactions on Image Processing, 2011, 20(5): 1401-1414

[4] Criminisi A, Cross G, Blake A, *et al*. Bilayer segmentation of live video[C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2006: 53-60

[5] Yin P, Criminisi A, Winn J, *et al*. Bilayer segmentation of webcam videos using tree-based classifiers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(1): 30-42

[6] Zhong F, Qin X Y, Peng Q S. Transductive segmentation of live video with non-stationary background[C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2010: 2189-2196

[7] Zhong F, Qin X Y, Peng Q S, *et al*. Discontinuity-aware video object cutout[J]. ACM Transactions on Graphics, 2012, 31(6): Article No.175

[8] Gordon G, Darrell T, Harville M, *et al*. Background estimation and removal based on range and color[C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 1999, 2: 2459-2464

[9] Camplani M, Salgado L. Background foreground segmentation with RGB-D Kinect data: an efficient combination of classifiers[J]. Journal of Visual Communication and Image Representation, 2014, 25(1): 122-136

[10] Kolmogorov V, Criminisi A, Blake A, *et al*. Bi-layer segmentation of binocular stereo video[C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2005, 2: 407-414

[11] Crabb R, Tracey C, Puranik A, *et al*. Real-time foreground segmentation via range and color imaging[C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2008: 1-5

[12] Wang L, Gong M L, Zhang C X, *et al*. Automatic real-time video matting using time-of-flight camera and multichannel Poisson equations[J]. International Journal of Computer Vision, 2012, 97(1): 104-121

[13] Bai X, Wang J, Simons D, *et al*. Video SnapCut: robust video object cutout using localized classifiers[J]. ACM Transactions on Graphics, 2009, 28(3): Article No. 70

[14] Bai X, Sapiro G. A geodesic framework for fast interactive image and video segmentation and matting[C] //Proceedings of the 11th IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2007: 1-8

[15] Shi J, Tomasi C. Good features to track[C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 1994: 593-600

[16] Criminisi A, Sharp T, Blake A. GeoS: geodesic image segmentation[C] //Proceedings of the 10th European Conference on Computer Vision. Heidelberg: Springer, 2008: 99-112

[17] Criminisi A, Sharp T, Rother C, *et al*. Geodesic image and video editing[J]. ACM Transactions on Graphics, 2010, 29(5): Article No. 134

[18] Toivanen P J. New geodosic distance transforms for gray-scale images[J]. Pattern Recognition Letters, 1996, 17(5): 437-450

[19] Park J, Kim H, Tai Y W, *et al*. High quality depth map upsampling for 3D-TOF cameras[C] //Proceedings of IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2011: 1623-1630

[20] Min D B, Lu J B, Do M N. Depth video enhancement based on weighted mode filtering[J]. IEEE Transactions on Image Processing, 2012, 21(3): 1176-1190